Dear Examiner,

The attached search was run with the most recently released version of Compugen's search software, GenCore 5. With this update, several changes have occurred in the results of FrameSearches (protein query sequence vs nucleic acid databases or nucleic acid query sequence vs protein databases).

In reference to FrameSearches:

- The output format has been improved so that it more closely resembles the format for standard search output.
- Calculation of Percent Similarity has been changed for FrameSearches. The new method of calculation is more similar to the method used in NCBI's BLAST algorithm. The same results are found in the same order using GenCore 5 and the previous version of GenCore, but Percent Similarities are lower in GenCore 5 results.

  - The formula for % similarity calculation is:

$$100 * \frac{\text{matches + conservative substitutions}}{\text{alignment_length}}$$

  where "matches" is the number of identical matches and "conservative substitutions" is the number of non-identical positive matches.

  - GenCore 4.5 considers the match Thr vs GCT (Ala) to be a similarity since BLOSUM62 gives score of 0 to this match. It is marked by ':::' in the alignment:

```
Qy     46 AspSerThrAspAla.Met..Gly 52
          ||||||:::    ||| :::  |||
Db    605 GATTCCGCTGCTGCTAATTTTGGC 628
```

  GenCore 5 requires a positive score to consider a non-identical match a similarity, therefore the same 'match' is not emphasized in the new alignment:

```
Qy     46 AspSerThrAspAla.Met..Gly 52
          ||||||       ||| :::  |||
Db    605 GATTCCGCTGCTGCTAATTTTGGC 628
```

If you have any questions, please feel free to contact one of the searchers in Biotech/Chem Library.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 10:34:25 ; Search time 66 Seconds
(without alignments)
325.051 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	848	100.0	279	23	AAE18419	Escherichia coli s
2	845	99.6	279	23	AAE18426	Escherichia coli s
3	838	98.8	279	23	AAE18435	Escherichia coli s
4	837	98.7	279	23	AAE18436	Escherichia coli F
5	834	98.3	279	23	AAE18418	Escherichia coli s
6	834	98.3	279	23	AAE18422	Escherichia coli s
7	834	98.3	279	23	AAE18424	Escherichia coli s
8	834	98.3	279	23	AAE18433	Escherichia coli s
9	834	98.3	300	16	AAR76769	FimH protein deriv
10	834	98.3	300	16	AAR76745	FimH protein deriv
11	834	98.3	300	21	AAB16009	E. coli proliferat
12	834	98.3	300	21	AAY59456	E. coli PC31 FimH
13	834	98.3	300	23	AAU77488	Escherichia coli t
14	833	98.2	279	23	AAE18427	Escherichia coli s
15	832	98.1	279	23	AAE18417	Escherichia coli s
16	831	98.0	279	23	AAE18413	Escherichia coli s
17	831	98.0	279	23	AAE18432	Escherichia coli s
18	830	97.9	300	16	AAR76763	FimH protein deriv
19	829	97.8	279	22	AAB47074	Adhesin protein, F
20	829	97.8	279	22	AAY72515	Escherichia coli F
21	829	97.8	279	23	AAE18428	Escherichia coli s
22	829	97.8	279	23	AAE18430	Escherichia coli s
23	829	97.8	296	22	AAB47073	Immunogenic comple
24	829	97.8	304	22	AAB47072	Immunogenic comple
25	828	97.6	300	16	AAR76768	FimH protein deriv
26	827	97.5	279	23	AAE18429	Escherichia coli s
27	827	97.5	280	23	AAE18420	Escherichia coli s
28	826	97.4	279	23	AAE18415	Escherichia coli s
29	826	97.4	279	23	AAE18416	Escherichia coli s
30	826	97.4	279	23	AAE18434	Escherichia coli s
31	825	97.3	279	23	AAE18421	Escherichia coli s
32	825	97.3	300	16	AAR76771	FimH protein deriv
33	825	97.3	300	16	AAR76773	FimH protein deriv
34	825	97.3	300	16	AAR76774	FimH protein deriv
35	824	97.2	300	16	AAR76772	FimH protein deriv
36	823	97.1	279	23	AAE18425	Escherichia coli s
37	822	96.9	300	16	AAR76770	FimH protein deriv
38	819	96.6	279	23	AAE18414	Escherichia coli s
39	819	96.6	279	23	AAE18423	Escherichia coli s
40	819	96.6	279	23	AAE18431	Escherichia coli s
41	818	96.5	300	16	AAR76775	FimH protein deriv
42	816	96.2	408	23	ABB09458	Fusion protein prt
43	815	96.1	300	16	AAR76766	FimH protein deriv
44	814	96.0	300	16	AAR76765	FimH protein deriv
45	814	96.0	300	16	AAR76764	FimH protein deriv

ALIGNMENTS

RESULT 1
AAE18419
ID AAE18419 standard; Protein; 279 AA.
XX
AC AAE18419;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain B240 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli B240.
XX
FH Key Location/Qualifiers
FT Misc-difference 201
FT /note= "Encoded by ACC"
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29358.

XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B240 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 100.0%; Score 848; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 2.1e-83;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
```

RESULT 2
AAE18426
ID AAE18426 standard; Protein; 279 AA.
XX
AC AAE18426;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC60 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli EC60.
XX
FH Key Location/Qualifiers
FT Misc-difference 27
FT /note= "Encoded by GCC"
FT Misc-difference 66
FT /note= "Encoded by AGC"
FT Misc-difference 70
FT /note= "Encoded by AGT"
FT Misc-difference 78
FT /note= "Encoded by AAT"
FT Misc-difference 119
FT /note= "Encoded by GCG"
FT Misc-difference 176
FT /note= "Encoded by CCT"
FT Misc-difference 201
FT /note= "Encoded by ACC"
FT Misc-difference 269
FT /note= "Encoded by CAA"
FT Misc-difference 273
FT /note= "Encoded by GGC"
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29365.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC60 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 99.6%; Score 845; DB 23; Length 279;
Best Local Similarity 99.4%; Pred. No. 4.4e-83;
Matches 160; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVVIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
```

RESULT 3
AAE18435
ID AAE18435 standard; Protein; 279 AA.
XX
AC AAE18435;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain G162 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli G162.
XX
FH Key Location/Qualifiers
FT Misc-difference 176
FT /note= "Encoded by CCT"
FT Misc-difference 201
FT /note= "Encoded by ACC"
XX
PN WO200204496-A2.
XX

PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29382.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Page 88-89; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain G162 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 98.8%; Score 838; DB 23; Length 279;
Best Local Similarity 98.1%; Pred. No. 2.5e-82;
Matches 158; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
```

RESULT 4
AAE18436
ID AAE18436 standard; Protein; 279 AA.
XX
AC AAE18436;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli FimH consensus protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli.
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli FimH consensus protein.
XX
SQ Sequence 279 AA;

Query Match 98.7%; Score 837; DB 23; Length 279;
Best Local Similarity 98.1%; Pred. No. 3.3e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
```

RESULT 5
AAE18418
ID AAE18418 standard; Protein; 279 AA.
XX
AC AAE18418;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain B238 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli B238.
XX
FH Key Location/Qualifiers
FT Misc-difference 24
FT /note= "Encoded by ATT"
FT Misc-difference 26
FT /note= "Encoded by CCC"
FT Misc-difference 176
FT /note= "Encoded by CCT"
FT Misc-difference 201
FT /note= "Encoded by ACC"
FT Misc-difference 273
FT /note= "Encoded by GCC"
FT Misc-difference 274
FT /note= "Encoded by GTG"
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.

XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29357.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B238 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 98.3%; Score 834; DB 23; Length 279;
Best Local Similarity 99.4%; Pred. No. 6.9e-82;
Matches 158; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWN 122
          ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGGVVIKAGSLIAVLILRQTNNYNSDDFQFVWN 147

Qy     123 IYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||
Db     148 IYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
```

RESULT 6
AAE18422
ID AAE18422 standard; Protein; 279 AA.
XX
AC AAE18422;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC42 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli EC42.
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29361.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC42 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 98.3%; Score 834; DB 23; Length 279;
Best Local Similarity 98.1%; Pred. No. 6.9e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVVIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 186
```

RESULT 7
AAE18424
ID AAE18424 standard; Protein; 279 AA.
XX
AC AAE18424;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC56 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli EC56.
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29363.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting

PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC56 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 98.3%; Score 834; DB 23; Length 279;
Best Local Similarity 98.1%; Pred. No. 6.9e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          ||||||||||||||||||||||||||||| ||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 186
```

RESULT 8
AAE18433
ID AAE18433 standard; Protein; 279 AA.
XX
AC AAE18433;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain J96 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli J96.
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29372.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Page 90-91; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain J96 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 98.3%; Score 834; DB 23; Length 279;
Best Local Similarity 98.1%; Pred. No. 6.9e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          ||||||||||||||||||||||||||||| ||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 186
```

RESULT 9
AAR76769
ID AAR76769 standard; protein; 300 AA.
XX
AC AAR76769;
XX
DT 15-MAR-1996 (first entry)
XX
DE FimH protein derived from E. coli clinical isolate CI#3.
XX
KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FimA; FimF; FimG; receptor binding site.
XX
OS Escherichia coli clinical isolate CI#3.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..300
FT /note= "Mature FimH"
XX
PN WO9520657-A1.
XX
PD 03-AUG-1995.
XX
PF 27-JAN-1995; 95WO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR WPI; 1995-275442/36.
DR N-PSDB; AAQ93071.
XX
PT Receptor specific bacterial adhesins - useful for targetting active
PT compounds and microbial cells to locations of receptors
XX
PS Example 1; Page 44-45; 152pp; English.
XX
CC The sequences given in AAR76763-76 are FimH proteins from various E.
CC coli clinical isolates. FimH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.

```
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             98.3%;  Score 834;  DB 16;  Length 300;
  Best Local Similarity   98.1%;  Pred. No. 7.6e-82;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVVIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 10
AAR76745
ID   AAR76745 standard; protein; 300 AA.
XX
AC   AAR76745;
XX
DT   13-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli K12 strain PC31.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli K12 strain PC31.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 88-89; 152pp; English.
XX
CC   This sequence represents the FimH protein from E. coli K12 strain
CC   PC31.  FimH is located at the tip of the  type 1 fimbriae and also
CC   intercalated at intervals in the fimbrial organelle.  Most forms of
CC   the FimH adhesin target, and bind to, oligosaccharide structures
CC   containing terminally located alpha-D-mannoside residues.  FimH
CC   contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  This
CC   sequence and those given in AAR76763-76 may be used in the production of
CC   a variant FimH adhesin which may be useful for targetting active
CC   compounds and microbial cells to locations comprising selected receptors
CC   to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             98.3%;  Score 834;  DB 16;  Length 300;
  Best Local Similarity   98.1%;  Pred. No. 7.6e-82;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 11
AAB16009
ID   AAB16009 standard; Protein; 300 AA.
XX
AC   AAB16009;
XX
DT   05-OCT-2000  (first entry)
XX
DE   E. coli proliferation associated protein sequence SEQ ID NO:367.
XX
KW   Escherichia coli; E. coli; proliferation; inhibition; screening;
KW   antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
OS   Escherichia coli.
XX
PN   WO200044906-A2.
XX
PD   03-AUG-2000.
XX
PF   27-JAN-2000; 2000WO-US02200.
XX
PR   27-JAN-1999;   99US-0117405.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zyskind J,  Ohlsen KL,  Trawick J,  Forsyth RA,  Froelich JM,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2000-514822/46.
DR   N-PSDB; AAA66015.
XX
PT   Novel polynucleotides and polypeptides associated with microorganism
```

PT proliferation, used to identify inhibitors of bacterial growth and
PT proliferation, for use in antisense therapy -
XX
PS Claim 11; Page 274-275; 316pp; English.
XX
CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC sequences derived from Escherichia coli which inhibit E. coli
CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC nucleotide and protein sequences associated with E. coli proliferation.
CC AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC proliferation inhibiting nucleotide inserts in an example from the
CC present invention. Methods from the present invention can be used to
CC identify a proliferation- required gene in a microorganism, by contacting
CC a microorganism with a proliferation-required gene activity inhibitory
CC nucleic acid identified in another organism, and determining if
CC inhibition occurs in the second microorganism. The nucleic acid sequences
CC identified as being required for bacterial growth and proliferation, can
CC be used for antisense therapy for killing bacteria.
XX
SQ Sequence 300 AA;

Query Match 98.3%; Score 834; DB 21; Length 300;
Best Local Similarity 98.1%; Pred. No. 7.6e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
Db 47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy 61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
Db 107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy 121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
Db 167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207

RESULT 12
AAY59456
ID AAY59456 standard; peptide; 300 AA.
XX
AC AAY59456;
XX
DT 29-MAR-2000 (first entry)
XX
DE E. coli PC31 FimH protein.
XX
KW Multifunctional adhesin protein; organic receptor; bioremediation;
KW biosorption; organic pollutant; herbicide; pesticide; toxic compound;
KW recycling; metal isolation; metal binding domain.
XX
OS Escherichia coli.
XX
PN WO9957276-A1.
XX
PD 11-NOV-1999.
XX
PF 21-APR-1999; 99WO-DK00223.
XX
PR 30-APR-1998; 98DK-0000598.
PR 01-MAY-1998; 98US-0083794.
XX
PA (GYRE-) GYRE LTD.
XX
PI Schembri MA, Klemm P;
XX
DR WPI; 2000-072233/06.
XX
PT Novel recombinant cells useful for bioremediation and recycling
PT processes -
XX
PS Claim 5; Page 5; 60pp; English.
XX
CC This sequence is the E. coli FimH protein, which is an adhesin protein.
CC The invention relates to a recombinant cell expressing a multifunctional
CC adhesin (MA) protein on its surface. The MA protein has at least one
CC binding domain (BD1) capable of binding to an organic receptor, and at
CC least one binding domain (BD2) not naturally present in the adhesin, and
CC can bind to a compound to which the naturally occurring adhesin protein
CC does not substantially bind. Cells of the invention are used as
CC bioremediation or biosorption means to separate undesired compounds such
CC as organic pollutants including herbicides and pesticides, or toxic
CC compounds such as heavy metals from the environment, or for isolating
CC precious compounds such as metals for recycling purposes.
XX
SQ Sequence 300 AA;

Query Match 98.3%; Score 834; DB 21; Length 300;
Best Local Similarity 98.1%; Pred. No. 7.6e-82;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
Db 47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy 61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
Db 107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy 121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
Db 167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207

RESULT 13
AAU77488
ID AAU77488 standard; Protein; 300 AA.
XX
AC AAU77488;
XX
DT 05-JUN-2002 (first entry)
XX
DE Escherichia coli type I pilin protein, FimH.
XX
KW Immune response; primate; immunoglobulin; urogenital tract infection;
KW FimC-FimH complex; FimCH; IgG; human; urinary tract infection; UTI;
KW bladder infection; kidney infection; Enterobacteriaceae; bactourea;
KW pregnant woman; diabetic; immunocompromised; HIV; cancer;
KW human immunodeficiency virus infection; end stage renal disease;
KW type I pilin protein; adhesin; FimH.
XX
OS Escherichia coli isolate J96.
XX
PN WO200215928-A1.
XX
PD 28-FEB-2002.
XX
PF 28-NOV-2000; 2000WO-US32398.
XX
PR 18-AUG-2000; 2000US-226146P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Ballou WR;
XX
DR WPI; 2002-280859/32.
DR N-PSDB; ABK11187.
XX
PT Stimulating immune response in a primate for preventing, treating
PT bacterial induced diseases such as diseases of urinary tract, by
PT administering bacterial adhesive proteins, preferably FimC-FimH
PT polypeptide complex -
XX
PS Claim 8; Page 89-90; 92pp; English.
XX

```
CC   The present invention relates to a method of inducing an immune response
CC   in a primate. The response involves immunoglobulin (Ig) molecules that
CC   bind a bacterial adhesin protein, preferably an attachment domain of
CC   a type 1 pilin polypeptide (e.g. FimH) associated with a bacterium
CC   causing urogenital tract infections (e.g. Escherichia coli). The method
CC   comprises administering a purified FimH polypeptide, a FimC-FimH (FimCH)
CC   complex, or immunogenic fragments of these. The method is useful for
CC   inducing IgG molecules in a primate, especially human, to reduce or
CC   prevent the incidence of urogenital tract infections, particularly
CC   urinary tract infection (UTI), bladder infection, or kidney infection,
CC   caused by a bacterium of the family Enterobacteriaceae, preferably
CC   E. coli. The method can be used in a human subject that has suffered
CC   more than two urogenital infections within one year, has asymptomatic
CC   bactourea, is a pregnant woman or a diabetic, is immunocompromised, has
CC   a human immunodeficiency virus (HIV) infection, has cancer, is in
CC   remission from cancer, or is at risk for end stage renal disease. The
CC   method is useful for vaccinating a primate against urogenital tract
CC   infections, for treating or ameliorating the symptoms of urogenital
CC   tract infections, and also for slowing or preventing progression of
CC   a urinary tract infection into end stage renal disease. The present
CC   sequence represents E. coli FimH protein.
XX
SQ   Sequence   300 AA;

  Query Match             98.3%;  Score 834;  DB 23;  Length 300;
  Best Local Similarity   98.1%;  Pred. No. 7.6e-82;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||| |||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 14
AAE18427
ID   AAE18427 standard; Protein; 279 AA.
XX
AC   AAE18427;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC61 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC61.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 3
FT                   /note= "Encoded by TAT"
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29366.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC61 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             98.2%;  Score 833;  DB 23;  Length 279;
  Best Local Similarity   97.5%;  Pred. No. 8.8e-82;
  Matches  157;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVEYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 15
AAE18417
ID   AAE18417 standard; Protein; 279 AA.
XX
AC   AAE18417;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B228 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B228.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 203
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
```

```
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29356.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B228 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             98.1%;  Score 832;  DB 23;  Length 279;
  Best Local Similarity   97.5%;  Pred. No. 1.1e-81;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||||| ||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSAHDVTVTLPDYRGSVPIPLTVY 186


RESULT 16
AAE18413
ID   AAE18413 standard; Protein; 279 AA.
XX
AC   AAE18413;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B210 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B210.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 141
FT                   /note= "Encoded by GGT"
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 203
FT                   /note= "Encoded by GGT"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29352.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B210 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             98.0%;  Score 831;  DB 23;  Length 279;
  Best Local Similarity   97.5%;  Pred. No. 1.5e-81;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGIVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||| ||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDASARDVTVTLPDYRGSVPIPLTVY 186


RESULT 17
AAE18432
ID   AAE18432 standard; Protein; 279 AA.
XX
AC   AAE18432;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain G189 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli G189.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
```

PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29371.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain G189 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 98.0%; Score 831; DB 23; Length 279;
Best Local Similarity 97.5%; Pred. No. 1.5e-81;
Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||| ||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTKNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
```

RESULT 18
AAR76763
ID AAR76763 standard; protein; 300 AA.
XX
AC AAR76763;
XX
DT 15-MAR-1996 (first entry)
XX
DE FimH protein derived from E. coli clinical isolate KB21.
XX
KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FimA; FimF; FimG; receptor binding site.
XX
OS Escherichia coli clinical isolate KB21.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "Signal peptide"
FT Protein 22..300
FT /note= "Mature FimH"
XX
PN WO9520657-A1.
XX
PD 03-AUG-1995.
XX
PF 27-JAN-1995; 95WO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR WPI; 1995-275442/36.
XX
PT Receptor specific bacterial adhesins - useful for targetting active
PT compounds and microbial cells to locations of receptors
XX
PS Example 1; Page 44-45; 152pp; English.
XX
CC The sequences given in AAR76763-76 are FimH proteins from various E.
CC coli clinical isolates. FimH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FimH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimA and
CC the minor components FimF and FimG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of a variant FimH adhesin which
CC may be useful for targetting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.
XX
SQ Sequence 300 AA;

Query Match 97.9%; Score 830; DB 16; Length 300;
Best Local Similarity 97.5%; Pred. No. 2.1e-81;
Matches 157; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207
```

RESULT 19
AAB47074
ID AAB47074 standard; Protein; 279 AA.
XX
AC AAB47074;
XX
DT 08-MAY-2001 (first entry)
XX
DE Adhesin protein, FimH.
XX
KW Donor; pilus protein; pilin; adhesin; vaccine; urinary epithelia;
KW urinary tract infection; enterobacteriaceae.
XX
OS Escherichia coli.
XX
PN WO200104148-A2.
XX
PD 18-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-US19066.
XX
PR 13-JUL-1999; 99US-0143582.
PR 16-JUL-1999; 99US-0144359.
PR 23-FEB-2000; 2000US-0184442.

```
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Hultgren SJ,  Pinkner JS,  Sauer F,  Barnhart M,  Waksman G,  Knight S;
XX
DR   WPI; 2001-138315/14.
XX
PT   Immunogenic complexes and polypeptides for vaccinating against urinary
PT   tract disease, comprises a pilus protein component and a bacterial
PT   chaperone -
XX
PS   Disclosure; Fig 1; 92pp; English.
XX
CC   This sequence may be used as the pilus protein in the immunogenic
CC   complex of the invention.  The complex comprises a pilus protein
CC   component and a donor complement portion as part of the same amino
CC   acid sequence or as non-covalently linked fragments of a complex
CC   such that the correct conformation of the pilin is maintained.  The
CC   pilus protein component may be an adhesin or a pilin.  Pilus associated
CC   adhesins, such as FimH are relatively conserved proteins among
CC   different species and strains of bacteria, therefore vaccines
CC   incorporating the FimH antigen exhibit a broad spectrum of
CC   protection compared with current pilus-fiber based vaccines.  The
CC   immunogenic complexes act by disrupting pilus-mediated attachment
CC   of E. coli to urinary epithelia and may prevent or retard the
CC   development of urinary tract infections.  Vaccines containing the
CC   complexes are useful for preventing urinary tract disease in a human
CC   caused by the bacterium family enterobacteriaceae specifically
CC   Escherichia coli and may also be used in treating the disease.
XX
SQ   Sequence   279 AA;

  Query Match             97.8%;  Score 829;  DB 22;  Length 279;
  Best Local Similarity   97.5%;  Pred. No. 2.4e-81;
  Matches  157;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: |||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||:|||||||| ||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARNVTVTLPDYPGSVPIPLTVY 186


RESULT 20
AAY72515
ID   AAY72515 standard; Protein; 279 AA.
XX
AC   AAY72515;
XX
DT   02-MAY-2001  (first entry)
XX
DE   Escherichia coli FimH protein.
XX
KW   FimH; adhesin protein; type 1 pilus; mannose binding domain; MBD; COL;
KW   collagen binding domain; prophylaxis; therapy; urinary tract infection;
KW   UTI; immunogen; passive immunotherapy; vaccine; antibacterial.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..156
FT                   /label= Lectin_binding_domain
FT   Domain          1..20
FT                   /label= MBD-1
FT                   /note= "Mannose binding domain-1"
FT   Region          2..5
FT                   /note= "Beta strand 1"
FT   Region          9..11
FT                   /note= "Beta strand 2"
FT   Region          17..25
FT                   /note= "Beta strand 3"
FT   Binding-site    27
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    31
FT                   /label= FimC_chaperone_binding_site
FT   Region          32..37
FT                   /note= "Beta strand 4a"
FT   Region          38..41
FT                   /note= "3-10 helix"
FT   Region          42..46
FT                   /note= "Beta strand 4b"
FT   Domain          46..54
FT                   /label= MBD-2
FT                   /note= "Mannose binding domain 2"
FT   Domain          50..80
FT                   /label= COL
FT                   /note= "Collagen binding domain;  Forms a beta-sheet-
FT                   alpha-1-beta-sheet structure from strands 5, alpha-1
FT                   and 6"
FT   Region          54..63
FT                   /note= "Beta strand 5"
FT   Region          67..69
FT                   /note= "Alpha-1 helix"
FT   Region          71..77
FT                   /note= "Beta strand 6"
FT   Region          80..83
FT                   /note= "Beta strand 7"
FT   Region          89..95
FT                   /note= "Beta strand 8"
FT   Region          104..111
FT                   /note= "Beta strand 9"
FT   Binding-site    117
FT                   /label= FimC_chaperone_binding_site
FT   Region          125..135
FT                   /note= "Beta strand 10"
FT   Domain          127..148
FT                   /label= MBD-3
FT                   /note= "Mannose binding domain 3"
FT   Region          139..150
FT                   /note= "Beta strand 11"
FT   Binding-site    155
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    157..158
FT                   /label= FimC_chaperone_binding_site
FT   Domain          159..279
FT                   /label= Pilin_domain
FT   Binding-site    160
FT                   /label= FimC_chaperone_binding_site
FT   Region          161..165
FT                   /note= "Beta strand A'"
FT   Binding-site    162..163
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    168..171
FT                   /note= "FimC_chaperone binding site; Serves as beta
FT                   strand A''"
FT   Binding-site    173
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    181
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    183
FT                   /label= FimC_chaperone_binding_site
FT   Region          183..187
FT                   /note= "Beta strand B"
FT   Binding-site    191..192
FT                   /label= FimC_chaperone_binding_site
FT   Region          192..197
FT                   /note= "Beta strand C"
FT   Binding-site    196
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    198
```

```
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    215
FT                   /label= FimC_chaperone_binding_site
FT   Region          221..226
FT                   /note= "Beta strand D'"
FT   Region          235..239
FT                   /note= "Beta strand D''"
FT   Region          251..258
FT                   /note= "Beta strand E"
FT   Binding-site    266..279
FT                   /label= FimC_chaperone_binding_site
FT   Region          268..279
FT                   /note= "Beta strand F; contains conserved beta-
FT                   zipper motif"
XX
PN   WO200105978-A1.
XX
PD   25-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-US19402.
XX
PR   15-JUL-1999;   99US-0144016.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Hultgren SJ,  Langermann S;
XX
DR   WPI; 2001-159539/16.
XX
PT   Polypeptides useful as vaccines for prevention and/or treatment of
PT   diseases such as urinary tract infections, caused by
PT   Enterobacteriaceae, comprises mannose-binding domains derived from
PT   adhesin molecules -
XX
PS   Claim 1; Fig 3; 53pp; English.
XX
CC   The present sequence is FimH protein from Escherichia coli. FimH
CC   is an adhesin protein found in type 1 pili of bacteria of the
CC   family enterobacteriaceae, especially E. coli. The FimH
CC   protein comprises mannose-binding domains (MBDs) and collagen-
CC   binding domains (COL). The present invention relates to
CC   engineered polypeptides comprising one or more domains derived
CC   from FimH protein. These polypeptides are used to produce
CC   prophylactic vaccines which are useful for the prevention
CC   and/or treatment of diseases, such as urinary tract infection (UTI)
CC   caused by a bacterium of the family Enterobacteriaceae, especially
CC   E. coli in animals, in particular humans. They are useful as immunogens
CC   to stimulate the production of antibodies for use in passive immuno-
CC   therapy, as a diagnostic reagent and as a reagent in other processes
CC   such as affinity chromatography. The antibodies of the novel poly-
CC   peptides are also useful for research purposes for studying protein-
CC   lectin or collagen binding interactions.
XX
SQ   Sequence   279 AA;

  Query Match             97.8%;  Score 829;  DB 22;  Length 279;
  Best Local Similarity   97.5%;  Pred. No. 2.4e-81;
  Matches  157;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||:|||||||| ||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARNVTVTLPDYPGSVPIPLTVY 186


RESULT 21
AAE18428
ID   AAE18428 standard; Protein; 279 AA.
XX
AC   AAE18428;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC62 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC62.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29367.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC62 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             97.8%;  Score 829;  DB 23;  Length 279;
  Best Local Similarity   96.9%;  Pred. No. 2.4e-81;
  Matches  156;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSHFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: |||||:||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLMAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 22
AAE18430
```

```
ID   AAE18430 standard; Protein; 279 AA.
XX
AC   AAE18430;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC89 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC89.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 226
FT                   /note= "Encoded by GCG"
FT   Misc-difference 227
FT                   /note= "Encoded by CGC"
FT   Misc-difference 231
FT                   /note= "Encoded by GTT"
FT   Misc-difference 232
FT                   /note= "Encoded by ATT"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29369.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC89 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             97.8%;  Score 829;  DB 23;  Length 279;
  Best Local Similarity   96.9%;  Pred. No. 2.4e-81;
  Matches  156;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 23
AAB47073
ID   AAB47073 standard; Protein; 296 AA.
XX
AC   AAB47073;
XX
DT   08-MAY-2001  (first entry)
XX
DE   Immunogenic complex: FimH-linker-FimG N-terminal extension.
XX
KW   Donor; pilus protein; pilin; adhesin; vaccine; urinary epithelia;
KW   urinary tract infection; enterobacteriaceae.
XX
OS   Chimeric - Escherichia coli.
OS   Chimeric - Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..279
FT                   /label= FimH
FT                   /note= "Pilus protein component"
FT   Peptide         280..283
FT                   /note= "Linker"
FT   Peptide         284..296
FT                   /label= N-terminal extension of FimG
FT                   /note= "Donor strand component"
XX
PN   WO200104148-A2.
XX
PD   18-JAN-2001.
XX
PF   13-JUL-2000; 2000WO-US19066.
XX
PR   13-JUL-1999;   99US-0143582.
PR   16-JUL-1999;   99US-0144359.
PR   23-FEB-2000; 2000US-0184442.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Hultgren SJ,  Pinkner JS,  Sauer F,  Barnhart M,  Waksman G,  Knight S;
XX
DR   WPI; 2001-138315/14.
XX
PT   Immunogenic complexes and polypeptides for vaccinating against urinary
PT   tract disease, comprises a pilus protein component and a bacterial
PT   chaperone -
XX
PS   Claim 19; Page 81-82; 92pp; English.
XX
CC   This sequence represents the immunogenic complex of the invention.
CC   The complex comprises a pilus protein component and a donor complement
CC   portion as part of the same amino acid sequence attached through an
CC   amino acid linker. The linker is composed of a sequence which
CC   readily forms a loop such that the donor strand can loop back towards
CC   the pilus protein and form an anti-parallel structure.  Pilus associated
CC   adhesins, such as FimH are relatively conserved proteins among
CC   different species and strains of bacteria, therefore vaccines
CC   incorporating the FimH antigen exhibit a broad spectrum of
CC   protection compared with current pilus-fiber based vaccines.  The
CC   immunogenic complexes act by disrupting pilus-mediated attachment
CC   of E. coli to urinary epithelia and may prevent or retard the
CC   development of urinary tract infections.  Vaccines containing the
CC   complexes are useful for preventing urinary tract disease in a human
CC   caused by the bacterium family enterobacteriaceae specifically
CC   Escherichia coli and may also be used in treating the disease.
XX
SQ   Sequence   296 AA;

  Query Match             97.8%;  Score 829;  DB 22;  Length 296;
```

```
  Best Local Similarity   97.5%;  Pred. No. 2.6e-81;
  Matches  157;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||:|||||||| ||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARNVTVTLPDYPGSVPIPLTVY 186


RESULT 24
AAB47072
ID   AAB47072 standard; Protein; 304 AA.
XX
AC   AAB47072;
XX
DT   08-MAY-2001  (first entry)
XX
DE   Immunogenic complex: FimH-linker-G1 beta-strand of FimC.
XX
KW   Donor; pilus protein; pilin; adhesin; vaccine; urinary epithelia;
KW   urinary tract infection; enterobacteriaceae.
XX
OS   Chimeric - Escherichia coli.
OS   Chimeric - Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..279
FT                   /label= FimH
FT                   /note= "Pilus protein component"
FT   Peptide         280..289
FT                   /note= "Linker"
FT   Peptide         290..304
FT                   /label= G1 beta-strand of FimC
FT                   /note= "Donor strand component"
XX
PN   WO200104148-A2.
XX
PD   18-JAN-2001.
XX
PF   13-JUL-2000; 2000WO-US19066.
XX
PR   13-JUL-1999;   99US-0143582.
PR   16-JUL-1999;   99US-0144359.
PR   23-FEB-2000; 2000US-0184442.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Hultgren SJ,  Pinkner JS,  Sauer F,  Barnhart M,  Waksman G,  Knight S;
XX
DR   WPI; 2001-138315/14.
XX
PT   Immunogenic complexes and polypeptides for vaccinating against urinary
PT   tract disease, comprises a pilus protein component and a bacterial
PT   chaperone -
XX
PS   Claim 19; Page 80-81; 92pp; English.
XX
CC   This sequence represents the immunogenic complex of the invention.
CC   The complex comprises a pilus protein component and a donor complement
CC   portion as part of the same amino acid sequence attached through an
CC   amino acid linker. The linker is composed of a sequence which
CC   readily forms a loop such that the donor strand can loop back towards
CC   the pilus protein and form an anti-parallel structure.  Pilus associated
CC   adhesins, such as FimH are relatively conserved proteins among
CC   different species and strains of bacteria, therefore vaccines
CC   incorporating the FimH antigen exhibit a broad spectrum of
CC   protection compared with current pilus-fiber based vaccines.  The
CC   immunogenic complexes act by disrupting pilus-mediated attachment
CC   of E. coli to urinary epithelia and may prevent or retard the
CC   development of urinary tract infections.  Vaccines containing the
CC   complexes are useful for preventing urinary tract disease in a human
CC   caused by the bacterium family enterobacteriaceae specifically
CC   Escherichia coli and may also be used in treating the disease.
XX
SQ   Sequence   304 AA;

  Query Match             97.8%;  Score 829;  DB 22;  Length 304;
  Best Local Similarity   97.5%;  Pred. No. 2.7e-81;
  Matches  157;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||:|||||||| ||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARNVTVTLPDYPGSVPIPLTVY 186


RESULT 25
AAR76768
ID   AAR76768 standard; protein; 300 AA.
XX
AC   AAR76768;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CSH-50.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CSH-50.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
```

```
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             97.6%;  Score 828;  DB 16;  Length 300;
  Best Local Similarity   97.5%;  Pred. No. 3.4e-81;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTRQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: |||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||||||||||||||| |||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 26
AAE18429
ID   AAE18429 standard; Protein; 279 AA.
XX
AC   AAE18429;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC80 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC80.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 27
FT                   /note= "Encoded by GTC"
FT   Misc-difference 165
FT                   /note= "Encoded by GCT"
FT   Misc-difference 171
FT                   /note= "Encoded by ACT"
FT   Misc-difference 269
FT                   /note= "Encoded by CAA"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29368.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC80 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             97.5%;  Score 827;  DB 23;  Length 279;
  Best Local Similarity   96.9%;  Pred. No. 3.9e-81;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: |||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||| ||||| |||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSVRDVTVILPDYRGSVPIPLTVY 186


RESULT 27
AAE18420
ID   AAE18420 standard; Protein; 280 AA.
XX
AC   AAE18420;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B242 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B242.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 279..280
FT                   /note= "Encoded by CAA"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29359.
```

```
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B242 FimH protein.
XX
SQ   Sequence   280 AA;

  Query Match             97.5%;  Score 827;  DB 23;  Length 280;
  Best Local Similarity   96.9%;  Pred. No. 4e-81;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      26 PAVNVGQNLVVDLSTQTFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||| |||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSAHDVTVTLPDYRGSVPIPLTVY 186


RESULT 28
AAE18415
ID   AAE18415 standard; Protein; 279 AA.
XX
AC   AAE18415;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B217 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B217.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 62
FT                   /note= "Encoded by TCG"
FT   Misc-difference 70
FT                   /note= "Encoded by AAT"
FT   Misc-difference 78
FT                   /note= "Encoded by AGT"
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 234
FT                   /note= "Encoded by ACG"
FT   Misc-difference 244
FT                   /note= "Encoded by CGG"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29354.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B217 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             97.4%;  Score 826;  DB 23;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 5.1e-81;
  Matches  155;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||:|||||||:||||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 29
AAE18416
ID   AAE18416 standard; Protein; 279 AA.
XX
AC   AAE18416;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B223 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B223.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
```

```
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29355.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B223 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             97.4%;  Score 826;  DB 23;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 5.1e-81;
  Matches  155;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||:|||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: |||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 30
AAE18434
ID   AAE18434 standard; Protein; 279 AA.
XX
AC   AAE18434;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain NU14 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli NU14.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29373.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain NU14 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             97.4%;  Score 826;  DB 23;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 5.1e-81;
  Matches  155;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||:|||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: |||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 31
AAE18421
ID   AAE18421 standard; Protein; 279 AA.
XX
AC   AAE18421;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain DS17 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli DS17.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
```

```
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29360.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain DS17 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             97.3%;  Score 825;  DB 23;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 6.5e-81;
  Matches  155;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: |||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||| |||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDASARDVTVTLPDYRGSVPIPLTVY 186


RESULT 32
AAR76771
ID   AAR76771 standard; protein; 300 AA.
XX
AC   AAR76771;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate KS-54.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate KS-54.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93073.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             97.3%;  Score 825;  DB 16;  Length 300;
  Best Local Similarity   96.9%;  Pred. No. 7.2e-81;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: |||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||| |||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSAHDVTVTLPDYPGSVPIPLTVY 207


RESULT 33
AAR76773
ID   AAR76773 standard; protein; 300 AA.
XX
AC   AAR76773;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate MJ#9-3.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate MJ#9-3.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
```

```
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93067.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             97.3%;  Score 825;  DB 16;  Length 300;
  Best Local Similarity   96.9%;  Pred. No. 7.2e-81;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: |||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||||| ||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSAHDVTVTLPDYPGSVPIPLTVY 207


RESULT 34
AAR76774
ID   AAR76774 standard; protein; 300 AA.
XX
AC   AAR76774;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate MJ#31-3.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate MJ#31-3.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93068.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             97.3%;  Score 825;  DB 16;  Length 300;
  Best Local Similarity   96.9%;  Pred. No. 7.2e-81;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: |||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||||| ||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSAHDVTVTLPDYPGSVPIPLTVY 207


RESULT 35
AAR76772
ID   AAR76772 standard; protein; 300 AA.
XX
AC   AAR76772;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate U221-3.
```

```
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate U221-3.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93074.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             97.2%;  Score 824;  DB 16;  Length 300;
  Best Local Similarity   96.9%;  Pred. No. 9.2e-81;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||| ||||||||||: |||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVARYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 36
AAE18425
ID   AAE18425 standard; Protein; 279 AA.
XX
AC   AAE18425;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC58 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC58.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29364.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC58 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             97.1%;  Score 823;  DB 23;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 1.1e-80;
  Matches  155;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||||| |||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGSVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: |||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 37
AAR76770
ID   AAR76770 standard; protein; 300 AA.
```

```
XX
AC   AAR76770;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate F-18.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate F-18.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93063.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; .Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             96.9%;  Score 822;  DB 16;  Length 300;
  Best Local Similarity   96.3%;  Pred. No. 1.5e-80;
  Matches  155;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 38
AAE18414
ID   AAE18414 standard; Protein; 279 AA.
XX
AC   AAE18414;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B212 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B212.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29353.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B212 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             96.6%;  Score 819;  DB 23;  Length 279;
  Best Local Similarity   95.7%;  Pred. No. 2.9e-80;
  Matches  154;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||:||| |||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGIVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||| ||||||||||||||||||||||
```

```
Db    146 WNIYANNDVVVPTGGCDASARDVTVTLPDYRGSVPIPLTVY 186


RESULT 39
AAE18423
ID   AAE18423 standard; Protein; 279 AA.
XX
AC   AAE18423;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC45 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC45.
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29362.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC45 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             96.6%;  Score 819;  DB 23;  Length 279;
  Best Local Similarity   95.7%;  Pred. No. 2.9e-80;
  Matches  154;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||:|||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 186


RESULT 40
AAE18431
ID   AAE18431 standard; Protein; 279 AA.
XX
AC   AAE18431;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC95 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC95.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29370.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Page 87-88; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC95 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             96.6%;  Score 819;  DB 23;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 2.9e-80;
  Matches  155;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||:|||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTLVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 41
AAR76775
ID   AAR76775 standard; protein; 300 AA.
XX
AC   AAR76775;
```

```
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate MJ#11-2.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate MJ#11-2.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93064.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             96.5%;  Score 818;  DB 16;  Length 300;
  Best Local Similarity   95.7%;  Pred. No. 4.1e-80;
  Matches  154;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||| |||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDASARDVTVTLPDYPGSVPIPLTVY 207
```

```
RESULT 42
ABB09458
ID   ABB09458 standard; Protein; 408 AA.
XX
AC   ABB09458;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Fusion protein prtY-FimH-prtY.
XX
KW   Exoproteinase; immunostimulant; vaccine; anchor peptide;
KW   FimH; ciliated adhesive factor.
XX
OS   Lactobacillus helveticus.
OS   Escherichia coli.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..70
FT                   /label= signal_peptide
FT                   /note= "derived from L. helveticus"
FT   Protein         71..295
FT                   /label= FimH_fragment
FT                   /note= "derived from E. coli ciliated adhesive factor"
FT   Peptide         296..408
FT                   /label= anchor_peptide
FT                   /note= "derived from L. helveticus"
XX
PN   JP2002017357-A.
XX
PD   22-JAN-2002.
XX
PF   04-JUL-2000; 2000JP-0202442.
XX
PR   04-JUL-2000; 2000JP-0202442.
XX
PA   (CALV ) CALPIS SHOKUHIN KOGYO KK.
XX
DR   WPI; 2002-221706/28.
DR   N-PSDB; ABL52756.
XX
PT   New anchor peptide, useful for anchoring protein onto microbe -
XX
PS   Example 1; Page 10-11; 12pp; Japanese.
XX
CC   The invention relates to an anchor peptide derived from the
CC   exoproteinase of Lactobacillus helveticus FERM BP-6060.
CC   The activity of compositions of the invention may be described as
CC   immunostimulatory. The invention also includes a fusion protein,
CC   containing the anchor peptide and a signal peptide fused to a useful
CC   protein. The invention also includes a method for anchoring a useful
CC   protein onto the surface of a microbe. The microbe can be used as a
CC   vaccine. The current sequence represents a fusion protein referred to as
CC   prtY-FimH-prtY. This fusion protein contains a fragment of the FimH
CC   protein, which is a ciliated adhesive factor from E. coli, arranged
CC   between the the signal sequence and anchor sequences derived from L.
CC   helveticus exoproteinase (see ABB09456 and ABB09457).
XX
SQ   Sequence   408 AA;

  Query Match             96.2%;  Score 816;  DB 23;  Length 408;
  Best Local Similarity   95.7%;  Pred. No. 1.1e-79;
  Matches  154;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||| |||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      93 PAVNVGQKLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 152

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     153 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 212
```

```
Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          |||||||||||||||||||||||||||||| |||||||||
Db    213 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 253


RESULT 43
AAR76766
ID   AAR76766 standard; protein; 300 AA.
XX
AC   AAR76766;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CI#7.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CI#7.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93072.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             96.1%;  Score 815;  DB 16;  Length 300;
  Best Local Similarity   96.3%;  Pred. No. 8.7e-80;
  Matches  155;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          | |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PAVNVGQHLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||| |||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db    107 TTSETLRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          |||||||||||||||||||||||||||||| |||||||||
Db    167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 44
AAR76765
ID   AAR76765 standard; protein; 300 AA.
XX
AC   AAR76765;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CI#4.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CI#4.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93065.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
```

```
XX
SQ   Sequence   300 AA;

  Query Match             96.0%;  Score 814;  DB 16;  Length 300;
  Best Local Similarity   95.7%;  Pred. No. 1.1e-79;
  Matches  154;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||:|| ||||:|||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSETVKYNGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 45
AAR76764
ID   AAR76764 standard; protein; 300 AA.
XX
AC   AAR76764;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CI#12.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CI#12.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;    95WO-DK00042.
XX
PR   27-JAN-1994;    94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93075.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             96.0%;  Score 814;  DB 16;  Length 300;
  Best Local Similarity   95.7%;  Pred. No. 1.1e-79;
  Matches  154;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||| ||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSACDVTVTLPDYPGSVPIPLTVY 207


Search completed: November 28, 2002, 17:19:55
Job time : 68 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 10:36:15 ; Search time 24 Seconds
(without alignments)
278.237 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	834	98.3	300	1	FIMH_ECOLI	P08191 escherichia
2	384	45.3	304	1	YDEQ_ECOLI	P77588 escherichia
3	205	24.2	291	1	SFAH_ECOLI	P13431 escherichia
4	86.5	10.2	550	1	HEMA_IADH6	P12587 influenza a
5	85.5	10.1	315	1	HEMA_VACCI	P08714 vaccinia vi
6	85.5	10.1	550	1	HEMA_IADH3	P12584 influenza a
7	84.5	10.0	566	1	HEMA_IAZCO	P26139 influenza a
8	83.5	9.8	567	1	HEMA_IAVI7	P03435 influenza a
9	83.5	9.8	639	1	APRP_ENTHI	P20302 entamoeba h
10	83	9.8	519	1	LAC2_TRAVI	Q99046 trametes vi
11	82.5	9.7	315	1	HEMA_VACCT	P16561 vaccinia vi
12	81.5	9.6	550	1	HEMA_IAZH2	P11133 influenza a
13	80.5	9.5	550	1	HEMA_IADH2	P12583 influenza a
14	80	9.4	314	1	HEMA_VACCV	Q01218 vaccinia vi
15	80	9.4	385	1	NK3R_MOUSE	P47937 mus musculu
16	79	9.3	373	1	PLYD_ASPNG	P22864 aspergillus
17	78.5	9.3	550	1	HEMA_IADH5	P12586 influenza a
18	78.5	9.3	550	1	HEMA_IADH7	P12588 influenza a
19	77.5	9.1	249	1	HEMA_IAX31	P03438 influenza a
20	77.5	9.1	315	1	HEMA_VACCC	P20978 vaccinia vi
21	77.5	9.1	550	1	HEMA_IADH1	P12582 influenza a
22	77.5	9.1	550	1	HEMA_IADH4	P12585 influenza a
23	77.5	9.1	550	1	HEMA_IADHK	P43257 influenza a
24	77.5	9.1	550	1	HEMA_IAZH3	P11134 influenza a
25	77.5	9.1	566	1	HEMA_IAAIC	P03437 influenza a
26	76.5	9.0	550	1	HEMA_IADHL	P43258 influenza a
27	76.5	9.0	550	1	HEMA_IAGHK	P43260 influenza a
28	76.5	9.0	566	1	HEMA_IANT6	P03436 influenza a
29	76	9.0	371	1	YA0C_SCHPO	Q09680 schizosacch
30	76	9.0	558	1	ILVD_SULSO	Q97ub2 sulfolobus
31	76	9.0	928	1	PM11_CHLPN	O86164 chlamydia p
32	75.5	8.9	566	1	HEMA_IADA3	P26134 influenza a
33	75.5	8.9	566	1	HEMA_IAMAO	P26138 influenza a
34	75	8.8	519	1	LAC2_TRAVE	Q12718 trametes ve
35	74.5	8.8	375	1	PELC_ERWCH	P11073 erwinia chr
36	74.5	8.8	566	1	HEMA_IADU3	P03442 influenza a
37	74.5	8.8	2195	1	POLG_EC11G	P29813 e genome po
38	73.5	8.7	328	1	HEMA_IAEN6	P04664 influenza a
39	73	8.6	452	1	NK3R_RAT	P16177 rattus norv
40	73	8.6	591	1	FLGE_CAUCR	P35806 caulobacter
41	73	8.6	1656	1	OMPB_RICJA	O06653 r outer mem
42	72.5	8.5	383	1	GABB_HUMAN	Q06547 homo sapien
43	72.5	8.5	580	1	FLA2_PYRKO	Q9v2x0 pyrococcus
44	72	8.5	342	1	RUVB_RICPR	Q9zde5 rickettsia
45	72	8.5	550	1	HEMA_IADHM	P43259 influenza a

ALIGNMENTS

RESULT 1
FIMH_ECOLI
ID FIMH_ECOLI STANDARD; PRT; 300 AA.
AC P08191;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE FimH protein precursor.
GN FIMH OR B4320.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88038337; PubMed=2890081;
RA Klemm P., Christiansen G.;
RT "Three fim genes required for the regulation of length and mediation
RT of adhesion of Escherichia coli type 1 fimbriae.";
RL Mol. Gen. Genet. 208:439-445(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=90256291; PubMed=1971261;
RA Krogfelt K.A., Bergmans H., Klemm P.;
RT "Direct evidence that the FimH protein is the mannose-specific
RT adhesin of Escherichia coli type 1 fimbriae.";
RL Infect. Immun. 58:1995-1998(1990).
CC -!- FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
CC ADHESION OF TYPE 1 FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION
CC OF FIMBRIAE). ADHESIN RESPONSIBLE FOR THE BINDING TO D-MANNOSE. IT
CC IS LATERALLY POSITIONED AT INTERVALS IN THE STRUCTURE OF THE TYPE
CC 1 FIMBRIAE. IN ORDER TO INTEGRATE FIMH IN THE FIMBRIAE FIMF AND
CC FIMG ARE NEEDED.
CC -!- SIMILARITY: THIS PROTEIN EXHIBITS HOMOLOGY WITH THE FIMBRIAL
CC SUBUNIT PROTEIN FIMA.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; X05672; CAA29156.1; -.
DR EMBL; U14003; AAA97216.1; -.
DR EMBL; AE000502; AAC77276.1; -.

```
DR   PIR; S09563; S09563.
DR   EcoGene; EG10315; fimH.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
KW   Fimbria; Signal; Complete proteome.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    300       FIMH PROTEIN.
FT   CONFLICT    197    197       P -> R (IN REF. 1).
FT   CONFLICT    222    222       T -> H (IN REF. 1).
SQ   SEQUENCE   300 AA;  31473 MW;  939204A51658747D CRC64;

  Query Match             98.3%;  Score 834;  DB 1;  Length 300;
  Best Local Similarity   98.1%;  Pred. No. 1.3e-70;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 2
YDEQ_ECOLI
ID   YDEQ_ECOLI     STANDARD;      PRT;   304 AA.
AC   P77588;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical fimbrial-like protein ydeQ precursor.
GN   YDEQ OR B1502.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
CC   -!- SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000247; AAC74575.1; -.
DR   EMBL; D90792; BAA15175.1; -.
DR   EMBL; D90793; BAA15183.1; -.
DR   EcoGene; EG13799; ydeQ.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
KW   Hypothetical protein; Fimbria; Signal; Complete proteome.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    304       HYPOTHETICAL FIMBRIAL-LIKE PROTEIN YDEQ.
SQ   SEQUENCE   304 AA;  32069 MW;  8153C86E3087D99A CRC64;

  Query Match             45.3%;  Score 384;  DB 1;  Length 304;
  Best Local Similarity   47.9%;  Pred. No. 1.2e-28;
  Matches   78;  Conservative   30;  Mismatches   53;  Indels    2;  Gaps    2;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETI-TDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPF 59
           ||:  |||||||||  | | |||     ||:: | :|||: | | :: |:: ::  :|||
Db      48 PVIQPGQNLVVDLSQHISCWNDYGGWYDTDHINLVQGSAFAGSLQSYKGSLYWNNVTYPF 107

Qy      60 PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNS-DDFQ 118
           | |: |  :    :|  | |: ||:||| :|||:||||| :|| : : :     | :
Db     108 PLTTNTNVLDIGDKTPMPLPLKLYITPVGAAGGVVIKAGEVIARIHMYKIATLGSGNPRN 167

Qy     119 FVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           | ||| :||:||:||||| | :|:||| |||: ||  ||| ||
Db     168 FTWNIISNNNVVMPTGGCTVDSRNVTVDLPDFPGSAEIPLGVY 210


RESULT 3
SFAH_ECOLI
ID   SFAH_ECOLI     STANDARD;      PRT;   291 AA.
AC   P13431;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   S-fimbrial adhesin protein sfaH precursor.
GN   SFAH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:K15;
RX   MEDLINE=90158121; PubMed=2576095;
RA   Schmoll T., Hoschuetzky H., Morschhaeuser J., Lottspeich F.,
RA   Jann K., Hacker J.;
RT   "Analysis of genes coding for the sialic acid-binding adhesin and two
RT   other minor fimbrial subunits of the S-fimbrial adhesin determinant
RT   of Escherichia coli.";
RL   Mol. Microbiol. 3:1735-1744(1989).
CC   -!- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
CC       FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC       MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO
CC       COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.
CC   -!- FUNCTION: THIS PROTEIN HAS SOME INFLUENCE ON THE LEVEL OF
CC       FIMBRIATION OF THE CELL.
CC   -!- DISEASE: S-FIMBRIAL ADHESINS ENABLE PATHOGENIC E.COLI CAUSING
CC       URINARY-TRACT INFECTIONS OR NEWBORN MENINGITIS TO ATTACH TO
CC       GLYCOPROTEINS TERMINATING WITH ALPHA-SIALIC ACID-(2-3)-BETA-GAL.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16664; CAA34654.1; -.
DR   PIR; S15927; S15927.
DR   InterPro; IPR000259; Fimbrial.
```

```
DR   Pfam; PF00419; Fimbrial; 1.
KW   Fimbria; Signal.
FT   SIGNAL       1     23
FT   CHAIN       24    291       S-FIMBRIAL ADHESIN PROTEIN SFAH.
SQ   SEQUENCE   291 AA;  31297 MW;  E026AA4713C2F947 CRC64;

  Query Match             24.2%;  Score 205;  DB 1;  Length 291;
  Best Local Similarity   32.6%;  Pred. No. 5.6e-12;
  Matches   56;  Conservative   22;  Mismatches   82;  Indels   12;  Gaps    5;

Qy       1 PVVNVGQNL-VVDLSTQIFCHN-DYPETITDYVTLQRGSAYGGVLS-------NFSGTVK 51
           |||   ::: |:|||  : | | |      ||: : :|| :   |        :|:
Db      34 PVVEYDKSISVLDLSQLVSCQNEDSTGQNYDYLKILKGSGFSPALDTKTYGRLDFTSRPT 93

Qy      52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQ--T 109
             |     :     |     ||:|  ||| |     | ||  | |:| | : :  |
Db      94 GYARQLPLQFDLQVTEAFYQYGVWKPFPAKLYLYPEPGVFGKVINNGDLLATLYVNKFST 153

Qy     110 NNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGS-VPIPLTV 160
           : :  | |  || ||| : || | ||: :| | || | |  | :||||
Db     154 KGQEAGERNFTWRFYATNDVHIQTGTCRVSSNNVKVDLPSYPGGPVTVPLTV 205


RESULT 4
HEMA_IADH6
ID   HEMA_IADH6     STANDARD;      PRT;   550 AA.
AC   P12587; Q84017;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11362;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16742; AAA43148.1; -.
DR   PIR; F27813; HMIV98.
DR   HSSP; P03437; 1HGJ.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER      1      1
FT   CHAIN        1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN      330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD    22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     8      8       Y -> N (IN PIR DATA BANK).
SQ   SEQUENCE   550 AA;  61711 MW;  67BCD85F44736CFE CRC64;

  Query Match             10.2%;  Score 86.5;  DB 1;  Length 550;
  Best Local Similarity   25.0%;  Pred. No. 1.3;
  Matches   37;  Conservative   18;  Mismatches   58;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:          |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPNSGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |               : |:||   || :   || | |   ::| ::         |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :||  || |    |:    || |
Db     194 LYVQASGGVTVSTRRSQQTIIPNIGSRP 221


RESULT 5
HEMA_VACCI
ID   HEMA_VACCI     STANDARD;      PRT;   315 AA.
AC   P08714;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Hemagglutinin precursor.
GN   HA OR A56R.
OS   Vaccinia virus (strain IHD-J).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10251;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86181588; PubMed=3008418;
RA   Shida H.;
RT   "Nucleotide sequence of the vaccinia virus hemagglutinin gene.";
RL   Virology 150:451-462(1986).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: THE MAJOR 85 kDa GLYCOPROTEIN IS PHOSPHORYLATED AND SULFATED.
CC       O-LINKED GLYCOSYLATION IS REQUIRED FOR HEMAGGLUTINATION AND
CC       HEMADSORPTION ACTIVITIES OF INFECTED CELL MEMBRANES.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14783; AAA48251.1; -.
DR   PIR; A27789; HNVZVV.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
KW   Hemagglutinin; Envelope protein; Late protein; Glycoprotein; Signal;
KW   Transmembrane; Immunoglobulin domain; Phosphorylation; Sulfation.
FT   SIGNAL       1     16
FT   CHAIN       17    315       HEMAGGLUTININ.
FT   DOMAIN      17    279       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   280    303       POTENTIAL.
FT   DOMAIN     304    315       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      17    121       IG-LIKE V-TYPE DOMAIN.
FT   DISULFID    34    103       POTENTIAL.
FT   CARBOHYD    37     37       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    69     69       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   254    254       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
SQ   SEQUENCE   315 AA;  34834 MW;  25EE5C7A6EA19779 CRC64;

  Query Match             10.1%;  Score 85.5;  DB 1;  Length 315;
  Best Local Similarity   25.5%;  Pred. No. 0.85;
  Matches   36;  Conservative   17;  Mismatches   37;  Indels   51;  Gaps    7;

Qy      24 PETITDYVTLQRGSAYGGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNS 72
           || ||| |             : : || |          || |    || |||  : :
Db     171 PEPITDNV-----------EDHTDTVTYTSDSINTVSASSGES----TTDETPEPITDK 214

Qy      73 RTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVP 132
             |      :  | ||:: |:|              |    :||   ::: | :|| | |
Db     215 EEDHTVTDTVSYTTVSTSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPP 259

Qy     133 T--GGCDVSARDVTVTLPDYR 151
           |  ||       | :: :|:
Db     260 TTVGGS-------TTSISNYK 273


RESULT 6
HEMA_IADH3
ID   HEMA_IADH3     STANDARD;      PRT;   550 AA.
AC   P12584; Q84012; Q89793;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/33/80).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11359;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16739; AAA43145.1; -.
DR   PIR; C27813; HMIV33.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61577 MW;  6C30BF67CFDCB7DE CRC64;

  Query Match             10.1%;  Score 85.5;  DB 1;  Length 550;
  Best Local Similarity   25.0%;  Pred. No. 1.7;
  Matches   37;  Conservative   18;  Mismatches   58;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |         ::|  :  ||:         |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPASGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |              : |:||   || :   || | |   ::| ::        |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :||  || |    |:    || |
Db     194 LYVQASGGVTVSTRRSQQTIIPNIGSRP 221


RESULT 7
HEMA_IAZCO
ID   HEMA_IAZCO     STANDARD;      PRT;   566 AA.
AC   P26139;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Swine/Colorado/1/77).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11493;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92114135; PubMed=1731092;
RA   Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
RA   Webster R.G.;
RT   "Evolution of the H3 influenza virus hemagglutinin from human and
RT   nonhuman hosts.";
RL   J. Virol. 66:1129-1138(1992).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73774; -; NOT_ANNOTATED_CDS.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL        1     16
FT   CHAIN        17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       346    566       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    142    142       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   566 AA;  63295 MW;  07297D7B3197C4E5 CRC64;

  Query Match             10.0%;  Score 84.5;  DB 1;  Length 566;
```

```
  Best Local Similarity   26.4%;  Pred. No. 2.1;
  Matches   39;  Conservative   18;  Mismatches   56;  Indels   35;  Gaps    8;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |       |::|  :  |||         |   : ||
Db     112 NCYPYDVPDYASLRSLVASSGTLEFINEGFNWTGVTQNGGSS----ACKRGPDNGFFSRL 167

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY------ANND 128
           :  |   ||           |:||   |  :   || |||   ::| ::         |
Db     168 N--W---LY-----------KSGSTYPVQNVTMPNNDNSDKL-YIWGVHHPSTDKEQTD 209

Qy     129 VVVPTGG-CDVSARDVTVTLPDYRGSVP 155
           : |   |   || : :  |:    || |
Db     210 LYVQASGKVTVSTKRIQQTVIPNVGSRP 237


RESULT 8
HEMA_IAVI7
ID   HEMA_IAVI7     STANDARD;      PRT;   567 AA.
AC   P03435;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Victoria/3/75).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11483;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=80155186; PubMed=6153930;
RA   Min Jou W., Verhoeyen M., Devos R., Saman E., Fang R.,
RA   Huylebroeck D., Fiers W., Threlfall G., Barber C., Carey N.,
RA   Emtage S.;
RT   "Complete structure of the hemagglutinin gene from the human
RT   influenza A/Victoria/3/75 (H3N2) strain as determined from cloned
RT   DNA.";
RL   Cell 19:683-696(1980).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=80254693; PubMed=7402351;
RA   Verhoeyen M., Fang R., Jou W.M., Devos R., Huylebroeck D., Saman E.,
RA   Fiers W.;
RT   "Antigenic drift between the haemagglutinin of the Hong Kong
RT   influenza strains A/Aichi/2/68 and A/Victoria/3/75.";
RL   Nature 286:771-776(1980).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01098; CAA24281.1; -.
DR   EMBL; V01086; CAA24270.1; -.
DR   PIR; A04050; HMIVV.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL        1     16
FT   CHAIN        17    345       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       347    567       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     25     25       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     39     39       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     55     55       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     80     80       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    143    143       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    182    182       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    500    500       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   567 AA;  63422 MW;  824D98A880EC5DEF CRC64;

  Query Match             9.8%;  Score 83.5;  DB 1;  Length 567;
  Best Local Similarity   26.4%;  Pred. No. 2.6;
  Matches   39;  Conservative   17;  Mismatches   57;  Indels   35;  Gaps    7;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |       |::|  :  |||         |   : ||
Db     113 NCYPYDVPDYASLRSLVASSGTLEFINEGFNWTGVTQNGGSS----ACKRGPDSGFFSRL 168

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |   ||           |:||   |  :   || |||   ::| ::         |
Db     169 N--W---LY-----------KSGSTYPVQNVTMPNNDNSDKL-YIWGVHHPSTDKEQTN 210

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
           |  :|   || :   |:    || |
Db     211 LYVQASGKVTVSTKRSQQTIIPNVGSRP 238


RESULT 9
APRP_ENTHI
ID   APRP_ENTHI     STANDARD;      PRT;   639 AA.
AC   P20302;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Antigenic protein P1 (Pathogenic protein 1) (Fragment).
OS   Entamoeba histolytica.
OC   Eukaryota; Entamoebidae; Entamoeba.
OX   NCBI_TaxID=5759;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HM-1:IMSS;
RX   MEDLINE=89296955; PubMed=2544890;
RA   Tannich E., Horstmann R.D., Knobloch J., Arnold H.H.;
RT   "Genomic DNA differences between pathogenic and nonpathogenic
RT   Entamoeba histolytica.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:5118-5122(1989).
CC   -!- SIMILARITY: 88% IDENTITY WITH PROTEIN FROM NON-PATHOGENIC
CC       STRAIN.
DR   PIR; A32935; A32935.
KW   Antigen.
FT   NON_TER       1      1
SQ   SEQUENCE   639 AA;  72086 MW;  4752187AB956AEB5 CRC64;

  Query Match             9.8%;  Score 83.5;  DB 1;  Length 639;
  Best Local Similarity   23.6%;  Pred. No. 3.1;
  Matches   38;  Conservative   32;  Mismatches   58;  Indels   33;  Gaps    7;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSY-PF 59
           |: ::| :| :: :|| :  ||         | : || :||::   | |  :: | |  |
Db      28 PIDSIGFDLGLN-TTQPYIIND---------TFKIGSPFGGMIYLRSDTT-FTNSFYVTF 76

Qy      60 PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQF 119
             | : ||  |:: |   |      :| | | :  :    |:|   |  : :| |:
Db      77 SNVGRAPIINYNITTNEEWNSVL-----RNAPGNVAEIRTPGNRLVLTSRNIRSLEDAQY 131

Qy     120 VWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTV 160
           : : :              :|  :  ||| :      ||:|:
Db     132 ISDFWLK----------AISISNYAVTLEN------IPITL 156


RESULT 10
```

```
LAC2_TRAVI
ID   LAC2_TRAVI     STANDARD;      PRT;   519 AA.
AC   Q99046;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Laccase 2 precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
DE   (Urishiol oxidase) (Diphenol oxidase).
GN   LCC2.
OS   Trametes villosa (White-rot fungus).
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Aphyllophorales; Trametes.
OX   NCBI_TaxID=47662;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mycelium;
RX   MEDLINE=97076915; PubMed=8975613;
RA   Yaver D.S., Xu F., Golightly E.J., Brown K.M., Brown S.H.,
RA   Rey M.W., Schneider P., Halkier T., Mondorf K., Dalboge H.;
RT   "Purification, characterization, molecular cloning, and expression of
RT   two laccase genes from the white rot basidiomycete Trametes
RT   villosa.";
RL   Appl. Environ. Microbiol. 62:834-841(1996).
CC   -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC       PRODUCTS (PROBABLE).
CC   -!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
CC       H(2)O.
CC   -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC       CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC       3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC   -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L49377; AAC41687.1; -.
DR   HSSP; P37064; 1AOZ.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR002355; MultiCu_oxidse2.
DR   Pfam; PF00394; Cu-oxidase; 3.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
DR   PROSITE; PS00080; MULTICOPPER_OXIDASE2; FALSE_NEG.
KW   Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
KW   Lignin degradation; Multigene family.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    519       LACCASE 2.
FT   DOMAIN       22    147       PLASTOCYANIN-LIKE 1.
FT   DOMAIN      159    301       PLASTOCYANIN-LIKE 2.
FT   DOMAIN      368    490       PLASTOCYANIN-LIKE 3.
FT   METAL        84     84       COPPER (TYPE 2) (BY SIMILARITY).
FT   METAL        86     86       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL       129    129       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL       131    131       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL       415    415       COPPER (TYPE 1) (BY SIMILARITY).
FT   METAL       418    418       COPPER (TYPE 2) (BY SIMILARITY).
FT   METAL       420    420       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL       472    472       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL       473    473       COPPER (TYPE 1) (BY SIMILARITY).
FT   METAL       474    474       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL       478    478       COPPER (TYPE 1) (BY SIMILARITY).
FT   METAL         ?      ?       COPPER (TYPE 1) (BY SIMILARITY).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    228    228       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    361    361       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    456    456       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   519 AA;  55777 MW;  DBB4EA526F684740 CRC64;

  Query Match             9.8%;  Score 83;  DB 1;  Length 519;
  Best Local Similarity   28.5%;  Pred. No. 2.7;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db     249 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 304

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||        |    |:|
Db     305 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 355

Qy     116 DFQFVWNIYANNDVVVP 132
           :|     : ||    |
Db     356 NF------FINNATFTP 366


RESULT 11
HEMA_VACCT
ID   HEMA_VACCT     STANDARD;      PRT;   315 AA.
AC   P16561;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Hemagglutinin precursor.
GN   HA OR A56R.
OS   Vaccinia virus (strain Tian Tan).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10253;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89328331; PubMed=2754392;
RA   Dongyan J., Zhiliang L., Qi J., Hao Y., Yunde H.;
RT   "Vaccinia virus hemagglutinin. A novel member of the immunoglobulin
RT   superfamily.";
RL   J. Exp. Med. 170:571-576(1989).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: THE MAJOR 85 kDa GLYCOPROTEIN IS PHOSPHORYLATED AND SULFATED.
CC       O-LINKED GLYCOSYLATION IS REQUIRED FOR HEMAGGLUTINATION AND
CC       HEMADSORPTION ACTIVITIES OF INFECTED CELL MEMBRANES.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15709; CAA33740.1; -.
DR   PIR; JL0108; HNVZVT.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
KW   Hemagglutinin; Envelope protein; Late protein; Glycoprotein; Signal;
KW   Transmembrane; Immunoglobulin domain; Phosphorylation; Sulfation.
FT   SIGNAL        1     16
FT   CHAIN        17    315       HEMAGGLUTININ.
FT   DOMAIN       17    279       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    280    303       POTENTIAL.
FT   DOMAIN      304    315       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       17    121       IG-LIKE V-TYPE DOMAIN.
FT   DISULFID     34    103       POTENTIAL.
FT   CARBOHYD     37     37       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD     69     69       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    254    254       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   315 AA;  34772 MW;  E049448640879FF4 CRC64;

  Query Match             9.7%;  Score 82.5;  DB 1;  Length 315;
  Best Local Similarity   23.8%;  Pred. No. 1.6;
  Matches   38;  Conservative   22;  Mismatches   43;  Indels   57;  Gaps    8;

Qy       4 NVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKY---------- 52
           |:  : | :::|        || ||| |             : : || |
Db     159 NLNCSSVFEIAT--------PEPITDNV-----------EDHTDTVTYTSDSINTVSAT 198

Qy      53 SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNY 112
           || |    || |||  : :   |      :  | ||:: |:|                |
Db     199 SGES----TTDETPEPITDKEEDHTVTDTVSYTTVSTSSGIV-------------TTKS 240

Qy     113 NSDDFQFVWNIYANNDVVVPTG-GCDVSARDVTVTLPDYR 151
             :||   ::: | :|| |  |  ||       | :: :|:
Db     241 TTDDAD-LYDTYNDNDTVPSTTVGCS------TTSISNYK 273


RESULT 12
HEMA_IAZH2
ID   HEMA_IAZH2     STANDARD;      PRT;   550 AA.
AC   P11133; Q84019; Q84020;
DT   01-JUL-1989 (Rel. 11, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE   chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11497;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88101364; PubMed=3336940;
RA   Kida H., Shortridge K.F., Webster R.G.;
RT   "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
RT   in China.";
RL   Virology 162:160-166(1988).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19057; AAA43212.1; -.
DR   PIR; B29971; HMIVS3.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Hemagglutinin; Envelope protein; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61437 MW;  1F2A7E758C531CE8 CRC64;

  Query Match             9.6%;  Score 81.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 3.9;
  Matches   36;  Conservative   19;  Mismatches   58;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:          |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPANGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |               : |:||   || :   || |||   ::| ::         |
Db     152 N--W---------------LTKSGSTYPVLNVTMPNNDNSDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || :    |:    || |
Db     194 LYVQASGRVTVSTKRSQQTMIPNAGSRP 221


RESULT 13
HEMA_IADH2
ID   HEMA_IADH2     STANDARD;      PRT;   550 AA.
AC   P12583; Q84011;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/8/80).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11358;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16738; AAA43144.1; -.
DR   PIR; B27813; HMIV80.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    137    137       K -> N (IN PIR DATA BANK).
SQ   SEQUENCE   550 AA;  61659 MW;  A107023ACC9CC353 CRC64;
```

```
  Query Match             9.5%;  Score 80.5;  DB 1;  Length 550;
  Best Local Similarity   23.9%;  Pred. No. 4.9;
  Matches   34;  Conservative   24;  Mismatches   61;  Indels   23;  Gaps    5;

Qy      21 NDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPV 80
           | ||  : || :|:        |   |||:|:    : :   ::       |:  |  |
Db      96 NCYPYDVPDYASLRS-------LVASSGTLKFITEGFTWTGVTQNG----GSKACKRGPA 144

Qy      81 ALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANNDVVVPT 133
           : : : ::     : |:||   || :   || | |   ::| ::         |  |  :
Db     145 SGFFSRLN----WLTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTNLYVQAS 199

Qy     134 GGCDVSARDVTVTLPDYRGSVP 155
           |   || |    |:   || |
Db     200 GRVTVSTRRSQQTIIPNIGSRP 221


RESULT 14
HEMA_VACCV
ID   HEMA_VACCV     STANDARD;      PRT;   314 AA.
AC   Q01218;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Hemagglutinin precursor.
GN   HA OR A56R OR SALG1R.
OS   Vaccinia virus (strain WR).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10254;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91259063; PubMed=2045793;
RA   Smith G.L., Chan Y.S., Howard S.T.;
RT   "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
RT   the right inverted terminal repeat.";
RL   J. Gen. Virol. 72:1349-1376(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92410621; PubMed=1529542;
RA   Cavallaro K.F., Esposito J.J.;
RT   "Sequences of the raccoon poxvirus hemagglutinin protein.";
RL   Virology 190:434-439(1992).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: THE MAJOR 85 kDa GLYCOPROTEIN IS PHOSPHORYLATED AND SULFATED.
CC       O-LINKED GLYCOSYLATION IS REQUIRED FOR HEMAGGLUTINATION AND
CC       HEMADSORPTION ACTIVITIES OF INFECTED CELL MEMBRANES.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D11079; BAA01829.1; -.
DR   EMBL; M93956; AAA48252.1; -.
DR   PIR; B43381; HNVZVW.
DR   PIR; JQ1793; JQ1793.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
KW   Hemagglutinin; Envelope protein; Late protein; Glycoprotein; Signal;
KW   Transmembrane; Immunoglobulin domain; Phosphorylation; Sulfation.
FT   SIGNAL        1     16
FT   CHAIN        17    314       HEMAGGLUTININ.
FT   DOMAIN       17    278       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    279    302       POTENTIAL.
FT   DOMAIN      303    314       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       17    121       IG-LIKE V-TYPE DOMAIN.
FT   DISULFID     34    103       POTENTIAL.
FT   CARBOHYD     37     37       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     69     69       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   314 AA;  34678 MW;  6F4322853244707F CRC64;

  Query Match             9.4%;  Score 80;  DB 1;  Length 314;
  Best Local Similarity   22.7%;  Pred. No. 2.8;
  Matches   42;  Conservative   21;  Mismatches   48;  Indels   74;  Gaps   10;

Qy       2 VVNVGQNLVVDLSTQIFCHN--------DY-------------PETITDYVTLQRGSAY 39
           :||     :|:      |:       ||              || ||| |
Db     127 IVNTDSESTIDIILSGSTHSPETSSKKPDYIDNSNCSSVFEIATPEPITDNV-------- 178

Qy      40 GGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
             : : || |          || |    || |||  :   : |      :  | ||
Db     179 ----EDHTDTVTYTSDSINTVSASSGES----TTDETPEPI-TDKEDHTVTDTVSYTTVS 229

Qy      89 SAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPT--GGCDVSARDVTVT 146
           :: |:|              |    :||   ::: | :|| | ||  ||        | :
Db     230 TSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPPTTVGGS-------TTS 267

Qy     147 LPDYR 151
           : :|:
Db     268 ISNYK 272


RESULT 15
NK3R_MOUSE
ID   NK3R_MOUSE     STANDARD;      PRT;   385 AA.
AC   P47937; Q61968;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Neuromedin K receptor (NKR) (Neurokinin B receptor) (NK-3 receptor)
DE   (NK-3R) (Fragment).
GN   TACR3 OR TAC3R.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Maroteaux L.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 103-328 FROM N.A.
RC   STRAIN=CBA; TISSUE=Brain;
RX   MEDLINE=94165478; PubMed=8120392;
RA   Cook G.A., Elliott D., Metwali A., Blum A.M., Sandor M., Lynch R.,
RA   Weinstock J.V.;
RT   "Molecular evidence that granuloma T lymphocytes in murine
RT   schistosomiasis mansoni express an authentic substance P (NK-1)
RT   receptor.";
RL   J. Immunol. 152:1830-1835(1994).
CC   -!- FUNCTION: THIS IS A RECEPTOR FOR THE TACHYKININ NEUROPEPTIDE
CC       NEUROMEDIN K (NEUROKININ B). IT IS ASSOCIATED WITH G PROTEINS
CC       THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER
CC       SYSTEM.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: THE ANCHORING OF THIS RECEPTOR TO THE PLASMA MEMBRANE IS
CC       PROBABLY MEDIATED BY THE PALMITOYLATION OF A CYSTEINE RESIDUE.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       HIGHEST TO OTHER TACHYKININS RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X87823; CAA61088.1; -.
DR   EMBL; L27827; AAA17893.1; -.
DR   HSSP; P02699; 1F88.
DR   MGD; MGI:892968; Tacr3.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation; Lipoprotein; Palmitate.
FT   DOMAIN       1     71       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     72    94       1 (POTENTIAL).
FT   DOMAIN       95    104      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     105   126      2 (POTENTIAL).
FT   DOMAIN       127   146      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     147   168      3 (POTENTIAL).
FT   DOMAIN       169   188      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     189   209      4 (POTENTIAL).
FT   DOMAIN       210   232      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     233   257      5 (POTENTIAL).
FT   DOMAIN       258   286      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     287   308      6 (POTENTIAL).
FT   DOMAIN       309   321      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     322   346      7 (POTENTIAL).
FT   DOMAIN       347   >385     CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     9     9        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     23    23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     40    40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     60    60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     145   220      BY SIMILARITY.
FT   LIPID        361   361      PALMITATE (POTENTIAL).
FT   CONFLICT     198   198      A -> S (IN REF. 2).
FT   CONFLICT     267   267      L -> P (IN REF. 2).
FT   NON_TER      385   385
SQ   SEQUENCE   385 AA;  43775 MW;  8A6974E0A455002C CRC64;

  Query Match             9.4%;  Score 80;  DB 1;  Length 385;
  Best Local Similarity   30.3%;  Pred. No. 3.5;
  Matches   30;  Conservative   15;  Mismatches   22;  Indels   32;  Gaps    6;

Qy      27 ITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK----PWPVAL 82
           ||::: || |        |||       |:   | ||: |  | :: |::     | :||
Db      30 ITEWLALQAG--------NFS-------SALGLPVTSQAPSQVRDNLTNQFVQPSWRIAL 74

Qy      83 YLTPVSSAGGLVIKA---GSLIAVLI------LRQTNNY 112
           :    | | |||:     |:|| : |      :|   ||
Db      75 W----SLAYGLVVAVAVFGNLIVIWIILAHKRMRTVTNY 109


RESULT 16
PLYD_ASPNG
ID   PLYD_ASPNG     STANDARD;      PRT;   373 AA.
AC   P22864;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Pectin lyase D precursor (EC 4.2.2.10) (PLD) (Pectin lyase I) (PLI).
GN   PELD.
OS   Aspergillus niger.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX   NCBI_TaxID=5061;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 20-30 AND 148-167.
RC   STRAIN=N756;
RX   MEDLINE=90323592; PubMed=2373363;
RA   Gysler C., Harmsen J.A.M., Kester H.C.M., Visser J., Heim J.;
RT   "Isolation and structure of the pectin lyase D-encoding gene from
RT   Aspergillus niger.";
RL   Gene 89:101-108(1990).
CC   -!- CATALYTIC ACTIVITY: Eliminative cleavage of pectin to give
CC       oligosaccharides with terminal 4-deoxy-6-methyl-alpha-D-galact-4-
CC       enuronosyl groups.
CC   -!- PTM: N-ACETYLGLUCOSAMINE WAS NOT FOUND IN PURIFIED PECTIN LYASE,
CC       BUT SOME O-GLYCOSYLATION MAY OCCUR.
CC   -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M55657; AAA32701.1; -.
DR   PIR; JH0155; JH0155.
DR   HSSP; Q01172; 1IDJ.
DR   InterPro; IPR002022; Amb_allergen.
DR   Pfam; PF00544; pec_lyase; 1.
KW   Lyase; Signal; Glycoprotein; Multigene family.
FT   SIGNAL       1     19
FT   CHAIN        20    373      PECTIN LYASE D.
FT   ACT_SITE     255   255      POTENTIAL.
FT   CARBOHYD     128   128      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     274   274      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     348   348      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     27    27       V -> E (IN REF. 1; AA SEQUENCE).
SQ   SEQUENCE   373 AA;  39038 MW;  B758DCCEFFE2345C CRC64;

  Query Match             9.3%;  Score 79;  DB 1;  Length 373;
  Best Local Similarity   23.6%;  Pred. No. 4.2;
  Matches   38;  Conservative   25;  Mismatches   72;  Indels   26;  Gaps    7;

Qy       9 LVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGV-LSNFSGTVKYSGSSYPFPTTSETPR 67
           |  |  :::   |:|    :||     |  |  | |   |  | :|| :| : |:   |:
Db     200 LGTDADSRVSITNNYINGESDYSATCDGHHYWNVYLDGSSDKVTFSG-NYLYKTSGRAPK 258

Qy      68 V-------VYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIA-VLILRQTNN-----YNS 114
           |       :||:  :     |  :     :|| |:  |:  : |  : :|:       ::|
Db     259 VQDNTYLHIYNNYWENNSGHAFEI----GSGGYVLAEGNYFSNVDTVLETDTFEGALFSS 314

Qy     115 DDFQFVWNIYANNDVVVPTGGCDVSARDVTV-------TLP 148
           |        |     |   | |::    ||        |||
Db     315 DSASSTCESYIGRSCVANVNGGDLTGTSTTVLSNLSGDTLP 355


RESULT 17
HEMA_IADH5
ID   HEMA_IADH5     STANDARD;      PRT;   550 AA.
AC   P12586; Q84015; Q84016;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11361;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
```

```
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16741; AAA43147.1; -.
DR   PIR; E27813; HMIV21.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      7      7       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    178    179       YV -> VI (IN PIR DATA BANK).
FT   CONFLICT    388    388       K -> T (IN PIR DATA BANK).
SQ   SEQUENCE   550 AA;  61856 MW;  48401C867A15BF8C CRC64;

  Query Match             9.3%;  Score 78.5;  DB 1;  Length 550;
  Best Local Similarity   25.0%;  Pred. No. 7.5;
  Matches   37;  Conservative   18;  Mismatches   58;  Indels   35;  Gaps    7;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  ::|| :|:   |  | |        ::|  :  ||:         |   : ||
Db      96 NCYPYDVSDYASLRSLVASSGTLEFLTEGFTWTGVTQNGGSN----ACKRGPASGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY------ANND 128
           :  |              : |:||   || :   || | |   :|| ::        |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YVWGVHHPSTNQEQTD 193

Qy     129 VVVPTGG-CDVSARDVTVTLPDYRGSVP 155
           : |   |   || |    |:    || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 18
HEMA_IADH7
ID   HEMA_IADH7     STANDARD;      PRT;   550 AA.
AC   P12588; Q84018; Q89470;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/10/85).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11363;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
```

```
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16743; AAA43149.1; -.
DR   PIR; G27813; HMIV15.
DR   HSSP; P03437; 3HMG.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61761 MW;  6EF81793281D53EB CRC64;

  Query Match             9.3%;  Score 78.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 7.5;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:         |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPNSGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |               : |:||   || :   || | |   ::| ::         |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
           |  :|   || |    |:    || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 19
HEMA_IAX31
ID   HEMA_IAX31     STANDARD;      PRT;   249 AA.
AC   P03438;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain]
DE   (Fragment).
GN   HA.
OS   Influenza A virus (strain X-31).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11489;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81030852; PubMed=7421990;
RA   Gething M.-J., Bye J., Skehel J.J., Waterfield M.;
RT   "Cloning and DNA sequence of double-stranded copies of haemagglutinin
RT   genes from H2 and H3 strains elucidates antigenic shift and drift in
RT   human influenza virus.";
RL   Nature 287:301-306(1980).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   PIR; A04051; HMIVH.
```

```
DR   PIR; A93233; A93233.
DR   HSSP; P03437; 1HGJ.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL        1     16
FT   CHAIN        17   >249       HEMAGGLUTININ HA1 CHAIN.
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     249    249
SQ   SEQUENCE   249 AA;  27373 MW;  9537AA970BB79183 CRC64;

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 249;
  Best Local Similarity   25.7%;  Pred. No. 3.6;
  Matches   44;  Conservative   21;  Mismatches   67;  Indels   39;  Gaps    9;

Qy       1 PVVNVGQNLVVDL---STQIFCHNDYPETITDYVTLQRGSAYGGVLS------NFSGTVK 51
           |  :| ||   ||    :: |  | ||  : || :|:   |  | |         ::|  :
Db      90 PHCDVFQNETWDLFVERSKAF-SNCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQ 148

Qy      52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNN 111
             ||:         |   : || :  |               : |:||   || :   ||
Db     149 NGGSN----ACKRGPGSGFFSRLN--W---------------LTKSGSTYPVLNVTMPNN 187

Qy     112 YNSDDFQFVWNIY---ANND----VVVPTGGCDVSARDVTVTLPDYRGSVP 155
             | |   ::| |:    | :      |  :|   || |    |:    || |
Db     188 DNFDKL-YIWGIHHPSTNQEQTSLYVQASGRVTVSTRRSQQTIIPNIGSRP 237


RESULT 20
HEMA_VACCC
ID   HEMA_VACCC     STANDARD;      PRT;   315 AA.
AC   P20978;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor.
GN   HA OR A56R.
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL   Virology 179:517-563(1990).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: THE MAJOR 85 kDa GLYCOPROTEIN IS PHOSPHORYLATED AND SULFATED.
CC       O-LINKED GLYCOSYLATION IS REQUIRED FOR HEMAGGLUTINATION AND
CC       HEMADSORPTION ACTIVITIES OF INFECTED CELL MEMBRANES.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35027; AAA48191.1; -.
DR   PIR; D42523; HNVZ4X.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
KW   Hemagglutinin; Envelope protein; Late protein; Glycoprotein; Signal;
KW   Transmembrane; Immunoglobulin domain; Phosphorylation; Sulfation.
FT   SIGNAL        1     16       BY SIMILARITY.
FT   CHAIN        17    315       HEMAGGLUTININ (BY SIMILARITY).
FT   DOMAIN       17    279       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    280    303       POTENTIAL.
FT   DOMAIN      304    315       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       17    121       IG-LIKE V-TYPE DOMAIN.
FT   DISULFID     34    103       POTENTIAL.
FT   CARBOHYD     37     37       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     69     69       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    254    254       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   315 AA;  34778 MW;  C7EE7D5EA12134D3 CRC64;

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 315;
  Best Local Similarity   21.3%;  Pred. No. 4.8;
  Matches   39;  Conservative   22;  Mismatches   53;  Indels   69;  Gaps    8;

Qy       2 VVNVGQNLVVDLSTQIFCHN--------DY--------------PETITDYVTLQRGSAY 39
           :||      :|:      |:        ||                || ||| |
Db     127 IVNTDSESTIDIILSGSTHSPETSSEKPDYIDNSNCSSVFEIATPEPITDNV-------- 178

Qy      40 GGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
                : : || |           || |    || |||  : :   |      :  | ||
Db     179 ----EDHTDTVTYTSDSINTVSATSGES----TTDETPEPITDKEEDHTVTDTVSYTTVS 230

Qy      89 SAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLP 148
           :: |:|              |    :||   ::: | :|| |  |     :    | ::
Db     231 TSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPST-----TVGSSTTSIS 270

Qy     149 DYR 151
           :|:
Db     271 NYK 273


RESULT 21
HEMA_IADH1
ID   HEMA_IADH1     STANDARD;      PRT;   550 AA.
AC   P12582; Q84021; Q84022;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/5/77).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11357;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16737; AAA43143.1; -.
DR   PIR; A27813; HMIV77.
DR   HSSP; P03437; 3HMG.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1     1
FT   CHAIN         1   328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330   550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8     8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22    22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38    38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165   165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285   285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483   483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61705 MW;  7E7ACFE716FC969A CRC64;

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 9.3;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:          |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPASGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |               : |:||   || :   || | |   ::| ::         |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTDQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || |    |:    || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 22
HEMA_IADH4
ID   HEMA_IADH4     STANDARD;      PRT;   550 AA.
AC   P12585; Q84013; Q84014;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/7/82).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11360;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M16740; AAA43146.1; -.
DR   PIR; D27813; HMIV89.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1     1
FT   CHAIN         1   328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330   550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8     8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22    22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38    38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165   165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285   285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483   483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61664 MW;  A16B2CF8CBBBD9D0 CRC64;

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 9.3;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:          |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPASGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |               : |:||   || :   || | |   ::| ::         |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || |    |:    || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 23
HEMA_IADHK
ID   HEMA_IADHK     STANDARD;      PRT;   550 AA.
AC   P43257;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hong Kong/7/75).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11364;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; D00929; BAA00769.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61549 MW;  864639B829FE1BA9 CRC64;

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 550;
  Best Local Similarity   24.7%;  Pred. No. 9.3;
  Matches   36;  Conservative   17;  Mismatches   58;  Indels   35;  Gaps    6;

Qy      23 YPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK 76
           ||  : || :|:   |  | |        ::|  :  |||         |   : || :
Db      98 YPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSS----ACKRGPASGFFSRLN- 152

Qy      77 PWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANNDV 129
            |                : |:||   || :   || | |   ::| ::        |
Db     153 -W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTNLY 195

Qy     130 VVPTGGCDVSARDVTVTLPDYRGSVP 155
           |  :|   || |    |:    || |
Db     196 VQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 24
HEMA_IAZH3
ID   HEMA_IAZH3     STANDARD;      PRT;   550 AA.
AC   P11134; Q84025; Q84026;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE   chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11498;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88101364; PubMed=3336940;
RA   Kida H., Shortridge K.F., Webster R.G.;
RT   "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
RT   in China.";
RL   Virology 162:160-166(1988).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19056; AAA43211.1; ALT_TERM.
DR   PIR; A29971; HMIVS2.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Hemagglutinin; Envelope protein; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61580 MW;  991F6D8BC02F24F2 CRC64;

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 550;
  Best Local Similarity   24.7%;  Pred. No. 9.3;
  Matches   36;  Conservative   17;  Mismatches   58;  Indels   35;  Gaps    6;

Qy      23 YPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK 76
           ||  : || :|:   |  | |        ::|  :  |||         |   : || :
Db      98 YPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSS----ACKRGPASGFFSRLN- 152

Qy      77 PWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANNDV 129
            |                : |:||   || :   || | |   ::| ::        |
Db     153 -W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTNLY 195

Qy     130 VVPTGGCDVSARDVTVTLPDYRGSVP 155
           |  :|   || |    |:    || |
Db     196 VQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 25
HEMA_IAAIC
ID   HEMA_IAAIC     STANDARD;      PRT;   566 AA.
AC   P03437;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Aichi/2/68).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=150147;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=80254693; PubMed=7402351;
RA   Verhoeyen M., Fang R., Min Jou W., Devos R., Huylebroeck D.,
RA   Saman E., Fiers W.;
RT   "Antigenic drift between the haemagglutinin of the Hong Kong
RT   influenza strains A/Aichi/2/68 and A/Victoria/3/75.";
RL   Nature 286:771-776(1980).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX   MEDLINE=81123029; PubMed=7464906;
RA   Wilson I.A., Skehel J.J., Wiley D.C.;
RT   "Structure of the haemagglutinin membrane glycoprotein of influenza
RT   virus at 3-A resolution.";
RL   Nature 289:366-373(1981).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE=88232903; PubMed=3374584;
RA   Weis W.I., Brown J.H., Cusack S.C., Paulson J.C., Skehel J.J.,
RA   Wiley D.C.;
RT   "Structure of the influenza virus haemagglutinin complexed with its
RT   receptor, sialic acid.";
RL   Nature 333:426-431(1988).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY OF A MUTANT WITH GLY-457.
RX   MEDLINE=90107940; PubMed=2295311;
RA   Weis W.I., Cusack S.C., Brown J.H., Daniels R.S., Skehel J.J.,
RA   Wiley D.C.;
```

```
RT   "The structure of a membrane fusion mutant of the influenza virus
RT   haemagglutinin.";
RL   EMBO J. 9:17-24(1990).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE=90230310; PubMed=2329580;
RA   Weis W.I., Bruenger A.T., Skehel J.J., Wiley D.C.;
RT   "Refinement of the influenza virus hemagglutinin by simulated
RT   annealing.";
RL   J. Mol. Biol. 212:737-761(1990).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE=94352388; PubMed=8072525;
RA   Bullough P.A., Hughson F.M., Skehel J.J., Wiley D.C.;
RT   "Structure of influenza haemagglutinin at the pH of membrane fusion.";
RL   Nature 371:37-43(1994).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).
RX   MEDLINE=98120975; PubMed=9461077;
RA   Fleury D., Wharton S.A., Skehel J.J., Knossow M., Bizebard T.;
RT   "Antigen distortion allows influenza virus to escape neutralization.";
RL   Nat. Struct. Biol. 5:119-123(1998).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J02090; AAA43178.1; -.
DR   EMBL; V01085; CAA24269.1; -.
DR   PIR; A93231; HMIVHA.
DR   PDB; 2HMG; 31-OCT-93.
DR   PDB; 3HMG; 31-OCT-93.
DR   PDB; 4HMG; 31-OCT-93.
DR   PDB; 5HMG; 31-JAN-94.
DR   PDB; 1HGD; 31-JAN-94.
DR   PDB; 1HGE; 31-JAN-94.
DR   PDB; 1HGF; 31-JAN-94.
DR   PDB; 1HGG; 31-JAN-94.
DR   PDB; 1HGH; 31-JAN-94.
DR   PDB; 1HGI; 31-JAN-94.
DR   PDB; 1HGJ; 31-JAN-94.
DR   PDB; 1HTM; 14-FEB-95.
DR   PDB; 2VIR; 29-APR-98.
DR   PDB; 2VIS; 29-APR-98.
DR   PDB; 2VIT; 29-APR-98.
DR   PDB; 2VIU; 29-APR-98.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT   SIGNAL       1    16
FT   CHAIN       17   344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN      346   566       HEMAGGLUTININ HA2 CHAIN.
FT   DISULFID    30   482       INTERCHAIN.
FT   DISULFID    68   293
FT   DISULFID    80    92
FT   DISULFID   113   155
FT   DISULFID   297   321
FT   DISULFID   489   493
FT   CARBOHYD    24    24       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    38    38       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    54    54       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    97    97       N-LINKED (GLCNAC. . .).
FT   CARBOHYD   181   181       N-LINKED (GLCNAC. . .).
FT   CARBOHYD   301   301       N-LINKED (GLCNAC. . .).
FT   CARBOHYD   499   499       N-LINKED (GLCNAC. . .).
FT   STRAND      27    34
FT   STRAND      40    42
FT   STRAND      50    52
FT   STRAND      55    57
FT   STRAND      59    60
FT   STRAND      67    70
FT   STRAND      74    76
FT   TURN        78    79
FT   HELIX       82    87
FT   TURN        88    88
FT   HELIX       90    95
FT   TURN        96    97
FT   STRAND      99    99
FT   STRAND     102   105
FT   TURN       107   108
FT   STRAND     116   117
FT   TURN       119   120
FT   HELIX      121   131
FT   STRAND     133   133
FT   STRAND     136   138
FT   TURN       144   145
FT   STRAND     146   150
FT   STRAND     152   157
FT   TURN       158   159
FT   STRAND     160   162
FT   TURN       165   166
FT   STRAND     167   173
FT   TURN       174   175
FT   STRAND     176   176
FT   STRAND     180   185
FT   STRAND     192   200
FT   HELIX      204   211
FT   STRAND     217   221
FT   STRAND     226   229
FT   STRAND     239   239
FT   TURN       240   241
FT   STRAND     242   242
FT   STRAND     245   253
FT   TURN       255   256
FT   STRAND     258   265
FT   STRAND     267   270
FT   STRAND     272   275
FT   STRAND     282   285
FT   STRAND     290   294
FT   STRAND     297   299
FT   TURN       300   301
FT   STRAND     302   304
FT   TURN       306   307
FT   STRAND     310   311
FT   STRAND     318   320
FT   STRAND     323   324
FT   STRAND     331   333
FT   STRAND     337   337
FT   TURN       347   348
FT   TURN       350   350
FT   STRAND     351   351
FT   TURN       352   354
FT   STRAND     355   355
FT   STRAND     359   359
FT   TURN       360   361
FT   STRAND     367   372
FT   TURN       374   375
FT   STRAND     378   382
FT   HELIX      383   401
FT   STRAND     406   407
FT   HELIX      421   474
FT   STRAND     475   477
FT   STRAND     482   485
FT   HELIX      491   498
FT   TURN       499   500
```

```
FT   HELIX       505    515
SQ   SEQUENCE   566 AA;  63415 MW;  E395659C23CAFECA CRC64;

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 566;
  Best Local Similarity   25.7%;  Pred. No. 9.6;
  Matches   44;  Conservative   21;  Mismatches   67;  Indels   39;  Gaps    9;

Qy       1 PVVNVGQNLVVDL---STQIFCHNDYPETITDYVTLQRGSAYGGVLS------NFSGTVK 51
           |  :| ||   ||    :: |  | ||  : || :|:   |  | |       ::| :
Db      90 PHCDVFQNETWDLFVERSKAF-SNCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQ 148

Qy      52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNN 111
             ||:         |   : || :  |               : |:||   || :   ||
Db     149 NGGSN----ACKRGPGSGFFSRLN--W--------------LTKSGSTYPVLNVTMPNN 187

Qy     112 YNSDDFQFVWNIY---ANND----VVVPTGGCDVSARDVTVTLPDYRGSVP 155
             | |   ::| |:    | :     |  :|   || |    |:    || |
Db     188 DNFDKL-YIWGIHHPSTNQEQTSLYVQASGRVTVSTRRSQQTIIPNIGSRP 237


RESULT 26
HEMA_IADHL
ID   HEMA_IADHL     STANDARD;      PRT;   550 AA.
AC   P43258;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hong Kong/64/76).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=45412;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00931; BAA00771.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61718 MW;  A351C56789E4BE9A CRC64;

  Query Match             9.0%;  Score 76.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 11;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
             | ||  : || :|:   |  | |         ::|  :  ||:          |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPANGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |                : |:||   || :   || | |   ::| ::          |
Db     152 N--W---------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || |    |:    || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 27
HEMA_IAGHK
ID   HEMA_IAGHK     STANDARD;      PRT;   550 AA.
AC   P43260;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Goose/Hong Kong/10/76).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=45414;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00930; BAA00770.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61676 MW;  9A1E094DA28BACD2 CRC64;

  Query Match             9.0%;  Score 76.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 11;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;
```

```
Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:          |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPANGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |              : |:||   || :   || | |   ::| ::         |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || |    |:    || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 28
HEMA_IANT6
ID   HEMA_IANT6     STANDARD;      PRT;   566 AA.
AC   P03436; Q84107;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/NT/60/68).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11446;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Subtype 29C;
RX   MEDLINE=81053698; PubMed=6253883;
RA   Both G.W., Sleigh M.J.;
RT   "Complete nucleotide sequence of the haemagglutinin gene from a human
RT   influenza virus of the Hong Kong subtype.";
RL   Nucleic Acids Res. 8:2561-2575(1980).
RN   [2]
RP   DISULFIDE BONDS.
RX   MEDLINE=80179105; PubMed=6768586;
RA   Dopheide T.A., Ward C.W.;
RT   "The disulphide bonds of a Hong Kong influenza virus hemagglutinin.";
RL   FEBS Lett. 110:181-183(1980).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- MISCELLANEOUS: STRAIN A/NT/60/68/29C IS A LABORATORY-ISOLATED
CC       VARIANT OF A/NT/60/68, AN EARLY STRAIN OF THE HONG KONG SUBTYPE.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01103; CAA24290.1; ALT_TERM.
DR   EMBL; V01103; CAA24291.1; ALT_INIT.
DR   PIR; A04051; HMIVH.
DR   HSSP; P03437; 1HGJ.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL        1     16
FT   CHAIN        17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       346    566       HEMAGGLUTININ HA2 CHAIN.
FT   DISULFID     30    482       INTERCHAIN.
FT   DISULFID     68    293
FT   DISULFID     80     92
FT   DISULFID    155    489
FT   DISULFID    297    321
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   566 AA;  63487 MW;  9A187F6A5E383671 CRC64;

  Query Match             9.0%;  Score 76.5;  DB 1;  Length 566;
  Best Local Similarity   25.1%;  Pred. No. 12;
  Matches   43;  Conservative   22;  Mismatches   67;  Indels   39;  Gaps    9;

Qy       1 PVVNVGQNLVVDL---STQIFCHNDYPETITDYVTLQRGSAYGGVLS------NFSGTVK 51
           |  :| ||   ||    :: |  | ||  : || :|:   |  | |        ::|  :
Db      90 PHCDVFQNETWDLFVERSKAF-SNCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQ 148

Qy      52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNN 111
           ||:          |   : || :  |               : |:||   || :   ||
Db     149 NGGSN----ACKRGPDSGFFSRLN--W--------------LTKSGSTYPVLNVTMPNN 187

Qy     112 YNSDDFQFVWNIY---ANND----VVVPTGGCDVSARDVTVTLPDYRGSVP 155
             | |   ::| ::    | :     |  :|   || |    |:    || |
Db     188 DNFDKL-YIWGVHHPSTNQEQTSLYVQASGRVTVSTRRSQQTIIPNIGSRP 237


RESULT 29
YA0C_SCHPO
ID   YA0C_SCHPO     STANDARD;      PRT;   371 AA.
AC   Q09680;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein C5H10.12c in chromosome I.
GN   SPAC5H10.12C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SIMILARITY: TO S.POMBE SPBC4C3.08 AND SPBC4C3.09.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z49811; CAA89962.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     17     33       POTENTIAL.
SQ   SEQUENCE   371 AA;  43688 MW;  890BAD4A0E66379B CRC64;

  Query Match             9.0%;  Score 76;  DB 1;  Length 371;
  Best Local Similarity   21.3%;  Pred. No. 8;
  Matches   33;  Conservative   27;  Mismatches   51;  Indels   44;  Gaps    6;

Qy       9 LVVDLSTQIFCHNDYPETITDYVTLQRGSAY----GGVLSNFSGTVKYSGSSYPFPTTSE 64
           |:: ::| :|  :|| :|: | :|  : :         ||    |   |  | :  |:| |
Db      24 LIIVMTTLVFNGHDYKQTLNDRLTSLKNNFVEENDNAVLKEEPGKYTYM-SLFTMPSTEE 82

Qy      65 ----------TPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNS 114
                     |: |:  |   :|                  | :| ||  : :
Db      83 DYYFNATRVLIHRLKYHPTTKSKYP--------------------IHILALRGVDEWKI 121

Qy     115 DDFQ------FVWNIYANNDVVVPTGGC--DVSAR 141
           : |:       | :  |::|:|  |     ::|||
Db     122 ERFRKDGASVIVIDPIASSDIVYDTSSFSQEISAR 156


RESULT 30
ILVD_SULSO
ID   ILVD_SULSO     STANDARD;      PRT;   558 AA.
AC   Q97UB2;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dihydroxy-acid dehydratase (EC 4.2.1.9) (DAD).
GN   ILVD OR SSO3107.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC   -!- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-
CC       oxobutanoate + H(2)O.
CC   -!- COFACTOR: Binds 1 4Fe-4S cluster (Potential).
CC   -!- PATHWAY: Valine and isoleucine biosynthesis; fourth step.
CC   -!- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE006901; AAK43209.1; -.
DR   InterPro; IPR004404; IlvD.
DR   InterPro; IPR000581; ILVD_EDD_family.
DR   Pfam; PF00920; ILVD_EDD; 1.
DR   ProDom; PD002691; ILVD_EDD_family; 1.
DR   TIGRFAMs; TIGR00110; ilvD; 1.
DR   PROSITE; PS00886; ILVD_EDD_1; 1.
DR   PROSITE; PS00887; ILVD_EDD_2; 1.
KW   Branched-chain amino acid biosynthesis; Lyase; Iron-sulfur; 4Fe-4S;
KW   Complete proteome.
FT   METAL       123    123       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       195    195       IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ   SEQUENCE   558 AA;  59421 MW;  059B2D16BAC62643 CRC64;

  Query Match             9.0%;  Score 76;  DB 1;  Length 558;
  Best Local Similarity   29.7%;  Pred. No. 13;
  Matches   30;  Conservative   10;  Mismatches   33;  Indels   28;  Gaps    4;

Qy      33 LQRGSAYGGVLSNFSGTVKYSGSSYPFPTT--SETPRVVYNSRTDKPWPVALYLTPVSSA 90
           |  |  :| ||:    |:| :   | :|    |   | | |              |:
Db     332 LDAGLLHGDVLTVTGKTMKQNLEQYKYPNVPHSHIVRDVKN--------------PIKPR 377

Qy      91 GGLVIKAGSLI---AVLILRQTN---------NYNSDDFQF 119
           ||:||  |||    ||: :  ||          |||:|  |
Db     378 GGIVILKGSLAPEGAVIKVAATNVVKFEGKAKVYNSEDDAF 418


RESULT 31
PM11_CHLPN
ID   PM11_CHLPN     STANDARD;      PRT;   928 AA.
AC   O86164; Q9K299;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable outer membrane protein pmp11 precursor (Polymorphic membrane
DE   protein 11) (Outer membrane protein 4).
GN   PMP11 OR OMP4 OR CPN0449 OR CP0302.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029/VR-1310;
RX   MEDLINE=99081766; PubMed=9864239;
RA   Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT   "Identification of two novel genes encoding 97- to 99-kilodalton outer
RT   membrane proteins of Chlamydia pneumoniae.";
RL   Infect. Immun. 67:375-383(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VR1310;
RX   MEDLINE=20007584; PubMed=10539856;
RA   Christiansen G., Boesen T., Hjerno K., Daugaard L., Mygind P.,
RA   Madsen A.S., Knudsen K., Falk E., Birkelund S.;
RT   "Molecular biology of Chlamydia pneumoniae surface proteins and their
RT   role in immunopathogenicity.";
RL   Am. Heart J. 138:S491-S495(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
```

```
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ001311; CAA04672.1; -.
DR   EMBL; AJ133034; CAB37072.1; -.
DR   EMBL; AE001628; AAD18593.1; -.
DR   EMBL; AE002192; AAF38159.1; ALT_INIT.
DR   EMBL; AP002546; BAA98658.1; -.
DR   PHCI-2DPAGE; O86164; -.
DR   TIGR; CP0302; -.
DR   InterPro; IPR003368; Chlamydia_PMP.
DR   InterPro; IPR003357; OMP.
DR   Pfam; PF02385; OMP; 1.
DR   Pfam; PF02415; DUF145; 2.
KW   Outer membrane; Signal; Multigene family; Complete proteome.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    928       PROBABLE OUTER MEMBRANE PROTEIN PMP11.
SQ   SEQUENCE   928 AA;  98903 MW;  788BCDD62C911402 CRC64;

  Query Match             9.0%;  Score 76;  DB 1;  Length 928;
  Best Local Similarity   23.8%;  Pred. No. 24;
  Matches   29;  Conservative   25;  Mismatches   34;  Indels   34;  Gaps    6;

Qy       3 VNVGQN-----LVVDLSTQIFCHNDYPE--TITDYVTLQRGSAYGGVLSNFSGTVKYSGS 55
           :|:| |     |    :  || ::      | :| : :  |||  | |: : ||: :||
Db     353 INIGSNGKFTELRAAKNHTIFFYDPITSEGTSSDVLKINNGSA--GALNPYQGTILFSGE 410

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVI--------------KAGSLI 101
           :    |::  :|  | ::       :  ||| :|| ::             :||||:
Db     411 T----LTADELKVADNLKSS------FTQPVSLSGGKLLLQKGVTLESTSFSQEAGSLL 459

Qy     102 AV 103
           :
Db     460 GM 461


RESULT 32
HEMA_IADA3
ID   HEMA_IADA3     STANDARD;      PRT;   566 AA.
AC   P26134;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Duck/Alberta/78/76).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11348;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92114135; PubMed=1731092;
RA   Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
RA   Webster R.G.;
RT   "Evolution of the H3 influenza virus hemagglutinin from human and
RT   nonhuman hosts.";
RL   J. Virol. 66:1129-1138(1992).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73771; -; NOT_ANNOTATED_CDS.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL        1     16
FT   CHAIN        17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       346    566       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   566 AA;  63534 MW;  FE19AB6FF9415B89 CRC64;

  Query Match             8.9%;  Score 75.5;  DB 1;  Length 566;
  Best Local Similarity   25.0%;  Pred. No. 15;
  Matches   37;  Conservative   17;  Mismatches   59;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |         ::|  :  ||:         |   : ||
Db     112 NCYPYDVPDYDSLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPASGFFSRL 167

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |               : |:||   || :   || | |   :|| ::         |
Db     168 N--W---------------LTKSGSTYPVLNVTMPNNDNFDKL-YVWGVHHPSTNQEQTN 209

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || |    |:    || |
Db     210 LYVQASGRVTVSTRRSQQTIIPNIGSRP 237


RESULT 33
HEMA_IAMAO
ID   HEMA_IAMAO     STANDARD;      PRT;   566 AA.
AC   P26138;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Mallard/New York/6874/78).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11436;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92114135; PubMed=1731092;
RA   Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
```

```
RA   Webster R.G.;
RT   "Evolution of the H3 influenza virus hemagglutinin from human and
RT   nonhuman hosts.";
RL   J. Virol. 66:1129-1138(1992).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73776; -; NOT_ANNOTATED_CDS.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL       1     16
FT   CHAIN       17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN      346    566       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD    24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   566 AA;  63574 MW;  BB206011C0BD9A3B CRC64;

  Query Match             8.9%;  Score 75.5;  DB 1;  Length 566;
  Best Local Similarity   25.0%;  Pred. No. 15;
  Matches   37;  Conservative   16;  Mismatches   60;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |         ::|  :  ||          |   : ||
Db     112 NCYPYDVPDYASLRSLVASSGTLEFITEAFTWTGVTQNGGSG----ACKRGPGNGFFSRL 167

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |                : |:||   || :   || | |   :|| ::        |
Db     168 N--W---------------LTKSGSAYPVLNVTMPNNDNFDKL-YVWGVHHPSTNQEQTN 209

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || |    |:    || |
Db     210 LYVQASGRVTVSTRRSQQTIIPNIGSRP 237


RESULT 34
LAC2_TRAVE
ID   LAC2_TRAVE     STANDARD;      PRT;   519 AA.
AC   Q12718; Q12716;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Laccase 2 precursor (EC 1.10.3.2) (Benzenediol:oxygen oxidoreductase)
DE   (Urishiol oxidase) (Diphenol oxidase) (Laccase I).
GN   LCC2 OR LCCI.
OS   Trametes versicolor (White-rot fungus).
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Aphyllophorales; Trametes.
OX   NCBI_TaxID=5325;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=52J;
RX   MEDLINE=97464057; PubMed=9322748;
RA   Ong E., Pollock W.B., Smith M.;
RT   "Cloning and sequence analysis of two laccase complementary DNAs from
RT   the ligninolytic basidiomycete Trametes versicolor.";
RL   Gene 196:113-119(1997).
CC   -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC       PRODUCTS (PROBABLE).
CC   -!- CATALYTIC ACTIVITY: 4 benzenediol + O(2) = 4 benzosemiquinone + 2
CC       H(2)O.
CC   -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC       CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC       3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC   -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U44851; AAA86659.1; -.
DR   EMBL; U44430; AAC49828.1; -.
DR   HSSP; P37064; 1AOZ.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR002355; MultiCu_oxidse2.
DR   Pfam; PF00394; Cu-oxidase; 3.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
DR   PROSITE; PS00080; MULTICOPPER_OXIDASE2; FALSE_NEG.
KW   Oxidoreductase; Signal; Copper; Metal-binding; Glycoprotein; Repeat;
KW   Lignin degradation; Multigene family.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21    519       LACCASE 2.
FT   DOMAIN      22    147       PLASTOCYANIN-LIKE 1.
FT   DOMAIN     159    301       PLASTOCYANIN-LIKE 2.
FT   DOMAIN     368    490       PLASTOCYANIN-LIKE 3.
FT   METAL       84     84       COPPER (TYPE 2) (BY SIMILARITY).
FT   METAL       86     86       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL      129    129       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL      131    131       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL      415    415       COPPER (TYPE 1) (BY SIMILARITY).
FT   METAL      418    418       COPPER (TYPE 2) (BY SIMILARITY).
FT   METAL      420    420       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL      472    472       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL      473    473       COPPER (TYPE 1) (BY SIMILARITY).
FT   METAL      474    474       COPPER (TYPE 3) (BY SIMILARITY).
FT   METAL      478    478       COPPER (TYPE 1) (BY SIMILARITY).
FT   METAL        ?      ?       COPPER (TYPE 1) (BY SIMILARITY).
FT   CARBOHYD    74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   228    228       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   361    361       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   456    456       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    69     69       D -> V (IN REF. 1; AAC49828).
SQ   SEQUENCE   519 AA;  55810 MW;  BFFB5B4CD0007702 CRC64;

  Query Match             8.8%;  Score 75;  DB 1;  Length 519;
  Best Local Similarity   27.7%;  Pred. No. 15;
  Matches   38;  Conservative   19;  Mismatches   50;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||     |        :|: :|  ::    :| |   |:| :     :| |:
Db     249 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 304

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|  |:    |:   |:|    | |   | | ||        |    |:|
Db     305 PVAEPTTTQTTSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 355

Qy     116 DFQFVWNIYANNDVVVP 132
           :|      : ||    |
```

```
Db    356 NF------FINNATFTP 366


RESULT 35
PELC_ERWCH
ID   PELC_ERWCH     STANDARD;      PRT;   375 AA.
AC   P11073;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Pectate lyase C precursor (EC 4.2.2.2).
GN   PELC.
OS   Erwinia chrysanthemi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pectobacterium.
OX   NCBI_TaxID=556;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EC16;
RX   MEDLINE=88298652; PubMed=3042750;
RA   Tamaki S.J., Gold S., Robeson M., Manulis S., Keen N.T.;
RT   "Structure and organization of the pel genes from Erwinia
RT   chrysanthemi EC16.";
RL   J. Bacteriol. 170:3468-3478(1988).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX   MEDLINE=93276270; PubMed=8502994;
RA   Yoder M.D., Keen N.T., Jurnak F.;
RT   "New domain motif: the structure of pectate lyase C, a secreted plant
RT   virulence factor.";
RL   Science 260:1503-1507(1993).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RA   Lietzke S.E., Scavetta R.D., Yoder M.D., Jurnak F.A.;
RT   "The refined three-dimensional structure of pectate lyase E from
RT   Erwinia chrysanthemi at 2.2-A resolution.";
RL   Plant Physiol. 111:73-92(1996).
CC   -!- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.
CC   -!- CATALYTIC ACTIVITY: Eliminative cleavage of pectate to give
CC       oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at
CC       their non-reducing ends.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1. PLADES
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19411; AAA24849.1; -.
DR   PIR; A31091; WZWC6C.
DR   PDB; 2PEC; 14-FEB-95.
DR   PDB; 1AIR; 16-JUN-97.
DR   InterPro; IPR002022; Amb_allergen.
DR   Pfam; PF00544; pec_lyase; 1.
KW   Lyase; Multigene family; Signal; 3D-structure.
FT   SIGNAL       1     22
FT   CHAIN       23    375       PECTATE LYASE C.
FT   DISULFID    94    177
FT   DISULFID   351    374
FT   ACT_SITE   240    240       POTENTIAL.
SQ   SEQUENCE   375 AA;  39943 MW;  F76DD8195A35B886 CRC64;

  Query Match             8.8%;  Score 74.5;  DB 1;  Length 375;
  Best Local Similarity   22.4%;  Pred. No. 11;
  Matches   37;  Conservative   23;  Mismatches   56;  Indels   49;  Gaps    8;

Qy     12 DLSTQIFCHNDYPETITDYVTLQRGS---AYGGVLSNFSGT---VKYSGSSYPFPTTSET 65
          |    |  |::|   :   : ||||    ||  : :| :|:   |: :| :

Db    222 DTGRNITYHHNYYNDVNARLPLQRGGLVHAYNNLYTNITGSGLNVRQNGQA--------- 272

Qy     66 PRVVYNSRTDKPWPVALYLTPVSSA------GGLVIKAGSLIAVLILRQTNNYNSDDFQF 119
            :: |:  :|       : ||:|       |  |:|  ::        |   :   :
Db    273 --LIENNWFEKA------INPVTSRYDGKNFGTWVLKGNNI--------TKPADFSTYSI 316

Qy    120 VWNI----YANNDVVVPTGG--------CDVSARDVTVTLPDYRG 152
           |      | | |    ||           |||: |   || | |
Db    317 TWTADTKPYVNADSWTSTGTFPTVAYNYSPVSAQCVKDKLPGYAG 361


RESULT 36
HEMA_IADU3
ID   HEMA_IADU3     STANDARD;      PRT;   566 AA.
AC   P03442;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Duck/Ukraine/1/63).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11374;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82025542; PubMed=6169439;
RA   Fang R., Min Jou W., Huylebroeck D., Devos R., Fiers W.;
RT   "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin
RT   gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
RT   hemagglutinin.";
RL   Cell 25:315-323(1981).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01087; CAA24271.1; -.
DR   PIR; A04053; HMIVDU.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL       1     16
FT   CHAIN       17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN      346    566       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD    24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   566 AA;  63530 MW;  E70F87F0AE1178F4 CRC64;

  Query Match             8.8%;  Score 74.5;  DB 1;  Length 566;
  Best Local Similarity   25.7%;  Pred. No. 18;
  Matches   44;  Conservative   18;  Mismatches   70;  Indels   39;  Gaps    8;

Qy      1 PVVNVGQNLVVDL---STQIFCHNDYPETITDYVTLQRGSAYGGVLS------NFSGTVK 51
          |  :| ||   ||    :  |  | ||  | || :|:   |  | |        ::|  :
```

```
Db     90 PHCDVFQNETWDLFVERSNAF-SNCYPYDIPDYASLRSLVASSGTLEFITEGFTWTGVTQ 148

Qy     52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNN 111
            |||         |   : || :  |                : |: |   || :   ||
Db    149 NGGSS----ACKRGPANGFFSRLN--W--------------LTKSESAYPVLNVTMPNN 187

Qy    112 YNSDDFQFVWNIY-------ANNDVVVPTGGCDVSARDVTVTLPDYRGSVP 155
           | |   ::| ::         |  |  :|   || |    |:    || |
Db    188 DNFDKL-YIWGVHHPSTNQEQTNLYVQASGRVTVSTRRSQQTIIPNIGSRP 237


RESULT 37
POLG_EC11G
ID   POLG_EC11G     STANDARD;      PRT;  2195 AA.
AC   P29813; Q66785;
DT   01-APR-1993 (Rel. 25, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
DE   (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
DE   (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
DE   P3A; Genome-linked protein VPG (P3B); Picornain 3C (EC 3.4.22.28)
DE   (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
OS   Echovirus 11 (strain Gregory).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Enterovirus.
OX   NCBI_TaxID=31705;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95282506; PubMed=7762294;
RA   Dahllund L., Nissinen L., Pulli T., Hyttinen V.P., Stanway G.,
RA   Hyypiae T.;
RT   "The genome of echovirus 11.";
RL   Virus Res. 35:215-222(1995).
RN   [2]
RP   SEQUENCE OF 822-2195 FROM N.A.
RX   MEDLINE=91011360; PubMed=2170575;
RA   Auvinen P., Hyypiae T.;
RT   "Echoviruses include genetically distinct serotypes.";
RL   J. Gen. Virol. 71:2133-2139(1990).
CC   -!- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
CC       AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
CC       PROTEASES.
CC   -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-|-Gly bond in the
CC       poliovirus polyprotein. In other picornavirus reactions Glu may be
CC       substituted for Gln, and Ser or Thr for Gly.
CC   -!- CATALYTIC ACTIVITY: Selective cleavage of Tyr-|-Gly bond in the
CC       picornavirus polyprotein. In other picornavirus reactions Glu may
CC       be substituted for Gln, and Ser or Thr for Gly.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC       CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS
CC       CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC   -!- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   -!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X80059; CAA56365.1; -.
DR   EMBL; D10582; BAA01439.1; -.
DR   PIR; A36642; GNNYEC.
DR   HSSP; P21404; 1D4M.
DR   MEROPS; C03.011; -.
DR   MEROPS; C03.022; -.
DR   InterPro; IPR004004; Calici_pol_hel.
DR   InterPro; IPR000199; Cys-protease-3C.
DR   InterPro; IPR003138; Pico_P1A.
DR   InterPro; IPR000081; Pico_P2A.
DR   InterPro; IPR002527; Pico_P2B.
DR   InterPro; IPR000605; RNA_helicase.
DR   InterPro; IPR001205; RNA_pol_P3D.
DR   InterPro; IPR001676; Rhv.
DR   Pfam; PF00073; rhv; 3.
DR   Pfam; PF00548; Cys-protease-3C; 1.
DR   Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR   Pfam; PF00910; RNA_helicase; 1.
DR   Pfam; PF00947; Pico_P2A; 1.
DR   Pfam; PF01552; Pico_P2B; 1.
DR   Pfam; PF02226; Pico_P1A; 1.
DR   PRINTS; PR00918; CALICIVIRUSNS.
DR   ProDom; PD001125; Cys-protease-3C; 1.
DR   ProDom; PD001274; Pico_P2B; 1.
DR   ProDom; PD001306; Pico_P2A; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase; Myristate;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT   CHAIN         2     69       COAT PROTEIN VP4.
FT   CHAIN        70    331       COAT PROTEIN VP2.
FT   CHAIN       332    569       COAT PROTEIN VP3.
FT   CHAIN       570    861       COAT PROTEIN VP1.
FT   CHAIN       862   1011       PICORNAIN 2A.
FT   CHAIN      1012   1110       CORE PROTEIN P2B.
FT   CHAIN      1111   1439       CORE PROTEIN P2C.
FT   CHAIN      1440   1528       CORE PROTEIN P3A.
FT   CHAIN      1529   1550       GENOME-LINKED PROTEIN VPG.
FT   CHAIN      1551   1733       PICORNAIN 3C.
FT   CHAIN      1734   2195       RNA-DIRECTED RNA POLYMERASE.
FT   LIPID         2      2       MYRISTATE (BY SIMILARITY).
FT   ACT_SITE   1697   1697       PROTEASE 3C (POTENTIAL).
FT   ACT_SITE   1711   1711       PROTEASE 3C (POTENTIAL).
FT   CONFLICT    823    827       RLCQY -> SYANT (IN REF. 2).
SQ   SEQUENCE   2195 AA;  245405 MW;  1CFC5DF288831AF0 CRC64;

  Query Match             8.8%;  Score 74.5;  DB 1;  Length 2195;
  Best Local Similarity   21.5%;  Pred. No. 92;
  Matches   39;  Conservative   24;  Mismatches   61;  Indels   57;  Gaps    9;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGT--VKYSGSSYP 58
            | :::|              | |      :    :   ||    | |   |::   | |
Db    753 PFISIG-------------NAYSNFYDGWSHFSQNGVYGYNTLNHMGQIYVRHVNGSSP 798

Qy     59 FPTTSETPRVVYNSRTDKPW-----------PVALYLTP------------------- 86
           | || | |: :  :  | |              :  ||
Db    799 LPMTS-TVRMYFKPKHVKAWVPRPPRLCQYKNASTVNFTPTNVTDKRTSINYIPETVKPD 857

Qy     87 VSSAGGLVIKAGSLIAVLILRQTNNY---NSDDFQFVWNIYANNDVVVPT---GGCDVSA 140
          :|: |    ::|: : |:  |  | :   ::|    ||  | | |::: |    |||| |
Db    858 LSNYGAFGYQSGA-VYVVNYRVVNRHLATHTDWQNCVWEDY-NRDLLISTTTAHGCDVIA 915

Qy    141 R 141
           |
Db    916 R 916


RESULT 38
HEMA_IAEN6
ID   HEMA_IAEN6     STANDARD;      PRT;   328 AA.
AC   P04664;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin [Contains: Hemagglutinin HA1 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/England/878/69).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11377;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81194918; PubMed=6164798;
RA   Sleigh M.J., Both G.W., Underwood P.A., Bender V.J.;
RT   "Antigenic drift in the hemagglutinin of the Hong Kong influenza
RT   subtype: correlation of amino acid changes with alterations in viral
RT   antigenicity.";
RL   J. Virol. 37:845-853(1981).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K03335; AAA43184.1; -.
DR   HSSP; P03437; 1HGE.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     328    328
SQ   SEQUENCE   328 AA;  36072 MW;  9C3A86B3A8D856FE CRC64;

  Query Match             8.7%;  Score 73.5;  DB 1;  Length 328;
  Best Local Similarity   24.3%;  Pred. No. 12;
  Matches   36;  Conservative   19;  Mismatches   58;  Indels   35;  Gaps    7;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:         |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPDSGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY---ANND--- 128
           :  |               : |:||   || :   || | |   ::| ::    | :
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTS 193

Qy     129 -VVVPTGGCDVSARDVTVTLPDYRGSVP 155
            |  :|   || |   |:   || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 39
NK3R_RAT
ID   NK3R_RAT       STANDARD;      PRT;   452 AA.
AC   P16177;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Neuromedin K receptor (NKR) (Neurokinin B receptor) (NK-3 receptor)
DE   (NK-3R).
GN   TACR3 OR TAC3R.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=90110113; PubMed=2153106;
RA   Shigemoto R., Yokota Y., Tsuchida K., Nakanishi S.;
RT   "Cloning and expression of a rat neuromedin K receptor cDNA.";
RL   J. Biol. Chem. 265:623-628(1990).
CC   -!- FUNCTION: THIS IS A RECEPTOR FOR THE TACHYKININ NEUROPEPTIDE
CC       NEUROMEDIN K (NEUROKININ B). IT IS ASSOCIATED WITH G PROTEINS
CC       THAT ACTIVATE A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER
CC       SYSTEM.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- PTM: THE ANCHORING OF THIS RECEPTOR TO THE PLASMA MEMBRANE IS
CC       PROBABLY MEDIATED BY THE PALMITOYLATION OF A CYSTEINE RESIDUE.
CC   -!- MISCELLANEOUS: THE RANK ORDER OF AFFINITY OF THIS RECEPTOR TO
CC       TACHYKININS IS: NEUROMEDIN K > SUBSTANCE K > SUBSTANCE P.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC       HIGHEST TO OTHER TACHYKININS RECEPTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J05189; AAA41688.1; -.
DR   PIR; A34916; A34916.
DR   HSSP; P02699; 1F88.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Phosphorylation; Lipoprotein; Palmitate.
FT   DOMAIN        1     71       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     72     94       1 (POTENTIAL).
FT   DOMAIN       95    104       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    105    126       2 (POTENTIAL).
FT   DOMAIN      127    146       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    147    168       3 (POTENTIAL).
FT   DOMAIN      169    188       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    189    209       4 (POTENTIAL).
FT   DOMAIN      210    232       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    233    257       5 (POTENTIAL).
FT   DOMAIN      258    286       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    287    308       6 (POTENTIAL).
FT   DOMAIN      309    321       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    322    346       7 (POTENTIAL).
FT   DOMAIN      347    452       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      9      9       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     40     40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    145    220       BY SIMILARITY.
FT   LIPID       361    361       PALMITATE (POTENTIAL).
SQ   SEQUENCE   452 AA;  51106 MW;  51D8EFC94F34CCAE CRC64;

  Query Match             8.6%;  Score 73;  DB 1;  Length 452;
  Best Local Similarity   28.3%;  Pred. No. 19;
  Matches   28;  Conservative   16;  Mismatches   23;  Indels   32;  Gaps    6;

Qy      27 ITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK----PWPVAL 82
           :|::: || |        |||       |:   | |:: |  |   : |::     | :||
Db      30 VTEWLALQAG--------NFS-------SALGLPATTQAPSQVRANLTNQFVQPSWRIAL 74

Qy      83 YLTPVSSAGGLVIKA---GSLIAVLI------LRQTNNY 112
           :    | | |||:     |:|| : |       :|   ||
Db      75 W----SLAYGLVVAVAVFGNLIVIWIILAHKRMRTVTNY 109


RESULT 40
FLGE_CAUCR
ID   FLGE_CAUCR     STANDARD;      PRT;   591 AA.
```

```
AC   P35806;
DT   01-JUN-1994 (Rel. 29, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Flagellar hook protein flgE.
GN   FLGE OR FLAK OR CC0902.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RA   Mullin D.A., Mullin A.H., Newton A.;
RT   "Organization and expression of Caulobacter crescentus genes needed
RT   for assembly and function of the flagellar hook.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN   [3]
RP   SEQUENCE OF 1-52 FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=86089096; PubMed=4078896;
RA   Ohta N., Chen L.S., Swanson E., Newton A.;
RT   "Transcriptional regulation of a periodically controlled flagellar
RT   gene operon in Caulobacter crescentus.";
RL   J. Mol. Biol. 186:107-115(1985).
CC   -!- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (Seé http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF072135; AAC33329.1; -.
DR   EMBL; AE005767; AAK22886.1; -.
DR   EMBL; M28551; AAA23042.1; -.
DR   TIGR; CC0902; -.
DR   InterPro; IPR001444; Flag_bb_rod.
DR   Pfam; PF00460; flg_bb_rod; 1.
DR   PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
KW   Flagella; Complete proteome.
SQ   SEQUENCE   591 AA;  61207 MW;  FD4EEA319479A161 CRC64;

  Query Match             8.6%;  Score 73;  DB 1;  Length 591;
  Best Local Similarity   22.1%;  Pred. No. 26;
  Matches   43;  Conservative   28;  Mismatches   72;  Indels   52;  Gaps    9;

Qy       8 NLVVDLS-TQIFCHNDYPE--TITDYVTLQRGSAYG----GVLSNFSGTV-------KYS 53
           || : :|    |   : ||  | ||  :  |  ::     | | | :|           |
Db      81 NLDISISGAGFFVTTEKPENLTATDTRSFTRAGSFQLDNLGYLRNDAGLYLQGWLADPVS 140

Qy      54 GSSYPFP----------------TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKA 97
           |   | |                 |  :| ||  |:       |||  ::       |   |
Db     141 GLITPDPSDLMQLASINVGSVGGTAEKTTRVGVNANLRSEQPVAAAVSYKVGTAGSPSKT 200

Qy      98 GSLIAVLILRQTNNYNSD--------------DFQFVWNIYANNDVVVPTG--GCDVSAR 141
                ::   ||::| |               : ::: :|   | ||:| ||    | :

Db     201 N-----VVDSATNSHNYDVVYSSTGIANPVSGNNEYLVDI-KENGVIVATGKVAYDAATN 254

Qy     142 DVTVTLPDYRGSVPI 156
           ::  :  ||:|: |:
Db     255 ELVSSTIDYKGASPV 269


RESULT 41
OMPB_RICJA
ID   OMPB_RICJA     STANDARD;      PRT;  1656 AA.
AC   O06653;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB.
OS   Rickettsia japonica.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=35790;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YH;
RA   Uchiyama T.;
RT   "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT   japonica.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC       SIMILARITY).
CC   -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY.
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB003681; BAA20138.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
KW   Antigen; S-layer; Cell wall.
FT   CHAIN         1   1338       120 KDA SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1339   1656        32 KDA BETA PEPTIDE.
FT   DOMAIN      528    533        POLY-GLY.
SQ   SEQUENCE   1656 AA;  168097 MW;  3132A69C9DD5999F CRC64;

  Query Match             8.6%;  Score 73;  DB 1;  Length 1656;
  Best Local Similarity   20.2%;  Pred. No. 91;
  Matches   35;  Conservative   24;  Mismatches   68;  Indels   46;  Gaps    7;

Qy       3 VNVGQNLVVDLS-TQIFCHNDYPETITDYVTLQRGSAYGGVLS--NFS----------GT 49
           :||  |   : | ||             |  :| |: : | |   ||:          |
Db    1194 INVQDNANANFSGTQ-----------TYTLIQGGARFNGTLGGPNFTVTGSNRFVNYGL 1241

Qy      50 VKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQT 109
           :: :   |    |:    :| |  |: |:           ||      ::   :    |
Db    1242 IRAANQDYVITRTNNAENIVTNDITNSPF-----------GGAPGVGQNVTTFVNATNT 1289

Qy     110 NNYN--------SDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
             ||        :|   ||  |  :    :     || |:|:   | :  |::
Db    1290 AAYNNLLLAKNSADSANFVGTIVTDTSAAITNAQLDV-AKDIQAQLGNRLGAL 1341
```

```
RESULT 42
GABB_HUMAN
ID   GABB_HUMAN     STANDARD;      PRT;   383 AA.
AC   Q06547;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   GA binding protein beta-1 chain (GABP-beta-1 subunit) (Transcription
DE   factor E4TF1-53) (GABPB1) (Nuclear respiratory factor-2 subunit beta
DE   2).
GN   GABPB1 OR GABPB OR E4TF1B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93180783; PubMed=8441384;
RA   Watanabe H., Sawada J.-I., Yano K.-I., Yamaguchi K., Goto M.,
RA   Handa H.;
RT   "cDNA cloning of transcription factor E4TF1 subunits with Ets and
RT   notch motifs.";
RL   Mol. Cell. Biol. 13:1385-1391(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95097980; PubMed=7799916;
RA   Gugneja S., Virbasius J.V., Scarpulla R.C.;
RT   "Four structurally distinct, non-DNA-binding subunits of human
RT   nuclear respiratory factor 2 share a conserved transcriptional
RT   activation domain.";
RL   Mol. Cell. Biol. 15:102-111(1995).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR CAPABLE OF INTERACTING WITH PURINE
CC       RICH REPEATS (GA REPEATS). NECESSARY FOR THE EXPRESSION OF THE
CC       ADENOVIRUS E4 GENE.
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; beta-1 (shown here) and beta-2
CC       (AC Q06545); are produced by alternative splicing.
CC   -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13317; BAA02574.1; -.
DR   EMBL; U13046; AAA65708.1; -.
DR   HSSP; Q00420; 1AWC.
DR   TRANSFAC; T01391; -.
DR   Genew; HGNC:4073; GABPB1.
DR   MIM; 600610; -.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; ank; 4.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 3.
DR   PROSITE; PS50088; ANK_REPEAT; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   Transcription regulation; Nuclear protein; Alternative splicing;
KW   ANK repeat; Repeat.
FT   REPEAT        5     34       ANK 1.
FT   REPEAT       37     66       ANK 2.
FT   REPEAT       70     99       ANK 3.
FT   REPEAT      103    132       ANK 4.
FT   REPEAT      136    166       ANK 5.
SQ   SEQUENCE   383 AA;  41321 MW;  07E7081A60016288 CRC64;

  Query Match             8.5%;  Score 72.5;  DB 1;  Length 383;
  Best Local Similarity   20.0%;  Pred. No. 18;
  Matches   36;  Conservative   38;  Mismatches   53;  Indels   53;  Gaps    9;

Qy       4 NVGQNLVVDLSTQIFCHNDYPETITDYV----------------------TLQRGSAYGG 41
           :: : | : :  ||  : : |:|:| :                        :| |::
Db     151 DLAEILQIAMQNQINTNPESPDTVTIHAATPQFIIGPGGVVNLTDETGVSAVQFGNSSTS 210

Qy      42 VLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGL--VIKAGS 99
           ||:  :  :  :| |  :|||| |    |::      :|  |  | :  |: :|
Db     211 VLATLAALAE---ASAPLSNSSETPVVA----TEE------VVTAESVDGAIQQVVSSGG 257

Qy     100 LIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLT 159
             : |:       :|  | : |:::     :|| |     : : ||:|| :  : :| |
Db     258 QQVITIV-------TDGIQ-LGNLHS-----IPTSGI---GQPIIVTMPDGQQVLTVPAT 301


RESULT 43
FLA2_PYRKO
ID   FLA2_PYRKO     STANDARD;      PRT;   580 AA.
AC   Q9V2X0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Flagellin B2 precursor.
GN   FLAB2.
OS   Pyrococcus kodakaraensis.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Thermococcus.
OX   NCBI_TaxID=69014;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KOD1;
RX   MEDLINE=99413236; PubMed=10483738;
RA   Nagahisa K., Ezaki S., Fujiwara S., Imanaka T., Takagi M.;
RT   "Sequence and transcriptional studies of five clustered flagellin
RT   genes from hyperthermophilic archaeon Pyrococcus kodakaraensis KOD1.";
RL   FEMS Microbiol. Lett. 178:183-190(1999).
CC   -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC       FORM THE FILAMENTS OF FLAGELLA.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEAL FLAGELLIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB018434; BAA84106.1; -.
DR   InterPro; IPR002774; Arch_flagellin.
DR   Pfam; PF01917; Arch_flagellin; 1.
KW   Flagella; Multigene family.
FT   PROPEP        1      9       BY SIMILARITY.
FT   CHAIN        10    580       FLAGELLIN B2.
SQ   SEQUENCE   580 AA;  61138 MW;  AC92D97B1CB96A2B CRC64;

  Query Match             8.5%;  Score 72.5;  DB 1;  Length 580;
  Best Local Similarity   25.4%;  Pred. No. 29;
  Matches   32;  Conservative   19;  Mismatches   42;  Indels   33;  Gaps    5;

Qy      42 VLSNFSGTVKYSGSSYPFPTTSETP------RVVYNSRTDKPWP-----VALYLTPVSSA 90
           || | || ::   ||    || |        :|:     | |       :|:|::| : :
Db      32 VLINTSGYLQQKASSTGRETTQEVASGLKIMKVIGYDPADPPASGKIERLAVYVSPNAGS 91

Qy      91 GGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTL--- 147
             |: :|   :|          :: | | ::| |      ||  |  ||    |: |
Db      92 SGIDMKKVRVI----------LSNGDKQAIYNYY------VPESGTFVSETPTTLKLAFA 135

Qy     148 ---PDY 150
              ||:
Db     136 TSEPDW 141


RESULT 44
RUVB_RICPR
```

```
ID   RUVB_RICPR     STANDARD;      PRT;   342 AA.
AC   Q9ZDE5;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Holliday junction DNA helicase ruvB.
GN   RUVB OR RP386.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   -!- FUNCTION: THE RUVA-RUVB COMPLEX IN THE PRESENCE OF ATP RENATURES
CC       CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE,
CC       INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN
CC       HOMOLOGOUS RECOMBINATION. RUVAB IS AN HELICASE THAT MEDIATES THE
CC       HOLLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND
CC       REANNELING (BY SIMILARITY).
CC   -!- FUNCTION: PARTICIPATES TO UV-TOLERANCE OF SYNECHOCYSTIS PCC 6803.
CC   -!- SUBUNIT: FORMS A COMPLEX WITH RUVA.
CC   -!- SIMILARITY: BELONGS TO THE RUVB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235271; CAA14843.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR004605; RuvB.
DR   Pfam; PF00004; AAA; 1.
DR   SMART; SM00382; AAA; 1.
DR   TIGRFAMs; TIGR00635; ruvB; 1.
KW   DNA repair; SOS response; ATP-binding; DNA recombination; Helicase;
KW   Complete proteome.
FT   NP_BIND      57     64       ATP (POTENTIAL).
SQ   SEQUENCE   342 AA;  38435 MW;  D2F9290E84079BA5 CRC64;

  Query Match             8.5%;  Score 72;  DB 1;  Length 342;
  Best Local Similarity   23.1%;  Pred. No. 17;
  Matches   25;  Conservative   15;  Mismatches   26;  Indels   42;  Gaps    4;

Qy      88 SSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANN----------------DVVV 131
           |:||  :|||  | ::|      |   :|  |:  |:  |                |:::
Db      80 STAGPAIIKAADLASIL-----TNLEKNDVLFIDEIHRLNTLVEEVLYSAMEDFELDIII 134

Qy     132 PTGGCDVSARDVTVTLPDY-----------------RGSVPIPLTVY 161
           |    :|| | :||| :                  |  :|: |  |
Db     135 GEGS---AARPVKITLPKFTLIGATTRFGLISNPLRDRFGIPMRLNFY 179


RESULT 45
HEMA_IADHM
ID   HEMA_IADHM     STANDARD;      PRT;   550 AA.
AC   P43259;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hong Kong/231/77).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=45411;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00932; BAA00772.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61762 MW;  6FEF8B49488C191A CRC64;

  Query Match             8.5%;  Score 72;  DB 1;  Length 550;
  Best Local Similarity   22.8%;  Pred. No. 30;
  Matches   36;  Conservative   18;  Mismatches   28;  Indels   76;  Gaps    8;

Qy      21 NDYPETITDYVTLQ------------RGSAYGGVLSN-------------FS--GTVK 51
           | ||  : || :|:              |  : ||  |             ||    :
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSNACKRGPANGFFSRLNWLT 155

Qy      52 YSGSSYP--------------------FPTTSE---------TPRVVYNSRT------- 74
           |||:||                    |:||:         : ||: ::|
Db     156 KSGSTYPVLNVTMPNNDNFDKLYIWGIHHPSTSQEQTTLYVQASGRVIVSTRRSQQTIIS 215

Qy      75 ---DKPW------PVALYLTPVSSAGGLVIKA-GSLIA 102
              :||       :::| | | |   ||| : |:|||
Db     216 NIGSRPWVRGQSGRISIYWTIVKSGDVLVINSNGNLIA 253


Search completed: November 28, 2002, 17:20:43
Job time : 28 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 10:37:10 ; Search time 26 Seconds
(without alignments)
1275.907 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	834	98.3	300	2	Q9S497	Q9s497 escherichia
2	834	98.3	300	2	Q9S496	Q9s496 escherichia
3	828	97.6	300	2	Q9S494	Q9s494 escherichia
4	828	97.6	300	2	Q9R5Y2	Q9r5y2 escherichia
5	827	97.5	300	2	Q9F5S1	Q9f5s1 escherichia
6	826	97.4	300	2	Q9F5S0	Q9f5s0 escherichia
7	824	97.2	300	16	Q8XBA6	Q8xba6 escherichia
8	823	97.1	300	2	Q9S495	Q9s495 escherichia
9	822	96.9	300	2	Q9F6Z7	Q9f6z7 escherichia
10	819	96.6	300	2	O87634	O87634 escherichia
11	819	96.6	303	2	Q9AP05	Q9ap05 escherichia
12	816	96.2	300	2	Q9S6R1	Q9s6r1 escherichia
13	756	89.2	243	2	Q08858	Q08858 klebsiella
14	618	72.9	166	2	Q9F5R9	Q9f5r9 escherichia
15	384	45.3	304	16	Q8XAX2	Q8xax2 escherichia
16	262	30.9	129	2	Q9XBV8	Q9xbv8 escherichia
17	207	24.4	299	2	Q53298	Q53298 escherichia
18	199.5	23.5	299	2	Q46686	Q46686 escherichia
19	89	10.5	684	16	Q9L1F8	Q9l1f8 streptomyce
20	88.5	10.4	315	12	Q9YN73	Q9yn73 vaccinia vi
21	87	10.3	270	12	Q9YN68	Q9yn68 cowpox viru
22	85.5	10.1	313	12	Q9YN66	Q9yn66 monkeypox v
23	85.5	10.1	313	12	Q8V4S6	Q8v4s6 monkeypox v
24	84	9.9	1755	17	O26812	O26812 methanobact
25	83.5	9.8	1114	5	Q05352	Q05352 entamoeba h
26	82.5	9.7	315	12	Q9JF47	Q9jf47 vaccinia vi
27	82.5	9.7	1155	17	O58048	O58048 pyrococcus
28	82	9.7	383	2	Q9X6L5	Q9x615 klebsiella
29	81.5	9.6	315	12	Q89120	Q89120 vaccinia vi
30	81.5	9.6	566	12	Q82496	Q82496 influenzavi
31	81	9.6	824	10	Q40713	Q40713 oryza sativ
32	80.5	9.5	473	10	Q9SE00	Q9se00 ipomoea bat
33	80.5	9.5	1304	2	Q9XC46	Q9xc46 rickettsia
34	80.5	9.5	1620	2	Q9KKB5	Q9kkb5 rickettsia
35	80	9.4	270	12	Q9YN72	Q9yn72 vaccinia vi
36	80	9.4	519	3	Q96UK8	Q96uk8 trametes ve
37	79.5	9.4	315	12	Q89182	Q89182 vaccinia vi
38	79.5	9.4	566	12	Q67126	Q67126 influenzavi
39	79.5	9.4	1924	12	Q9JFS2	Q9jfs2 ectromelia
40	79	9.3	421	9	O64302	O64302 bacteriopha
41	79	9.3	985	5	Q9VFW7	Q9vfw7 drosophila
42	79	9.3	1064	10	Q9FX19	Q9fx19 arabidopsis
43	79	9.3	3763	5	Q8T2A1	Q8t2a1 dictyosteli
44	78.5	9.3	398	16	Q8U9U1	Q8u9u1 agrobacteri
45	78.5	9.3	782	2	Q93SH4	Q93sh4 bradyrhizob

ALIGNMENTS

RESULT 1
Q9S497
ID Q9S497 PRELIMINARY; PRT; 300 AA.
AC Q9S497;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FimH.
GN FIMH.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=20327582; PubMed=10869080;
RA Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA Russell P.W., Orndorff P.E.;
RT "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT mutants and identification of a novel binding phenotype.";
RL J. Bacteriol. 182:4012-4021(2000).
DR EMBL; AF154925; AAD44319.1; -.
DR InterPro; IPR000259; Fimbrial.
DR Pfam; PF00419; Fimbrial; 1.
FT VARIANT 19 19 D -> A.
SQ SEQUENCE 300 AA; 31517 MW; D29340A1575A2358 CRC64;

Query Match 98.3%; Score 834; DB 2; Length 300;
Best Local Similarity 98.1%; Pred. No. 4e-71;
Matches 158; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166
```

```
Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          ||||||||||||||||||||||||||||| |||||||||
Db    167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 2
Q9S496
ID   Q9S496      PRELIMINARY;      PRT;   300 AA.
AC   Q9S496;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=20327582; PubMed=10869080;
RA   Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA   Russell P.W., Orndorff P.E.;
RT   "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT   mutants and identification of a novel binding phenotype.";
RL   J. Bacteriol. 182:4012-4021(2000).
DR   EMBL; AF154926; AAD44320.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT      21     21       L -> S.
SQ   SEQUENCE   300 AA;  31499 MW;  3ADF97A85F8CE478 CRC64;

  Query Match             98.3%;  Score 834;  DB 2;  Length 300;
  Best Local Similarity   98.1%;  Pred. No. 4e-71;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          ||||||||||||||||||||||||||||| |||||||||
Db    167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 3
Q9S494
ID   Q9S494      PRELIMINARY;      PRT;   300 AA.
AC   Q9S494;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=20327582; PubMed=10869080;
RA   Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA   Russell P.W., Orndorff P.E.;
RT   "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT   mutants and identification of a novel binding phenotype.";
RL   J. Bacteriol. 182:4012-4021(2000).
DR   EMBL; AF154928; AAD44322.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT      79     79       R -> L.
SQ   SEQUENCE   300 AA;  31516 MW;  B6073DF68366B86E CRC64;

  Query Match             97.6%;  Score 828;  DB 2;  Length 300;
  Best Local Similarity   97.5%;  Pred. No. 1.5e-70;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTRQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          ||||||||||||||||||||||||||||| |||||||||
Db    167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 4
Q9R5Y2
ID   Q9R5Y2      PRELIMINARY;      PRT;   300 AA.
AC   Q9R5Y2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   FIMH SUBUNIT=MANNOSE-sensitive type 1 fimbrial adhesin.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94131954; PubMed=7905476;
RA   Sokurenko E.V., Courtney H.S., Ohman D.E., Klemm P., Hasty D.L.;
RL   J. Bacteriol. 176:748-755(1994).
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   300 AA;  31552 MW;  E75328B7C366A319 CRC64;

  Query Match             97.6%;  Score 828;  DB 2;  Length 300;
  Best Local Similarity   97.5%;  Pred. No. 1.5e-70;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTRQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          ||||||||||||||||||||||||||||| |||||||||
Db    167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 5
Q9F5S1
ID   Q9F5S1      PRELIMINARY;      PRT;   300 AA.
AC   Q9F5S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
```

```
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=21289098; PubMed=11395476;
RA   Harris S.L., Spears P.A., Havell E.A., Hamrick T.S., Horton J.R.,
RA   Orndorff P.E.;
RT   "Characterization of Escherichia coli Type 1 Pilus Mutants with
RT   Altered Binding Specificities.";
RL   J. Bacteriol. 183:4099-4102(2001).
DR   EMBL; AF306535; AAG30925.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT      79     79       P -> R.
SQ   SEQUENCE   300 AA;  31457 MW;  9F7B0A35FF51F938 CRC64;

  Query Match             97.5%;  Score 827;  DB 2;  Length 300;
  Best Local Similarity   97.5%;  Pred. No. 1.8e-70;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTPQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 6
Q9F5S0
ID   Q9F5S0      PRELIMINARY;      PRT;   300 AA.
AC   Q9F5S0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=21289098; PubMed=11395476;
RA   Harris S.L., Spears P.A., Havell E.A., Hamrick T.S., Horton J.R.,
RA   Orndorff P.E.;
RT   "Characterization of Escherichia coli Type 1 Pilus Mutants with
RT   Altered Binding Specificities.";
RL   J. Bacteriol. 183:4099-4102(2001).
DR   EMBL; AF306536; AAG30926.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT     163    163       S -> F.
SQ   SEQUENCE   300 AA;  31413 MW;  922904051758746D CRC64;

  Query Match             97.4%;  Score 826;  DB 2;  Length 300;
  Best Local Similarity   97.5%;  Pred. No. 2.3e-70;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: ||||||||||||||||||||||||| |||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDSQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 7
Q8XBA6
ID   Q8XBA6      PRELIMINARY;      PRT;   300 AA.
AC   Q8XBA6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Minor fimbrial subunit, D-mannose specific adhesin.
GN   FIMH OR Z5918 OR ECS5279.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005662; AAG59502.1; -.
DR   EMBL; AP002569; BAB38702.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
KW   Complete proteome.
SQ   SEQUENCE   300 AA;  31459 MW;  68EFE713A3D372CB CRC64;

  Query Match             97.2%;  Score 824;  DB 16;  Length 300;
  Best Local Similarity   96.9%;  Pred. No. 3.5e-70;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: |||||||||||||| |||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTKNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 8
Q9S495
ID   Q9S495      PRELIMINARY;      PRT;   300 AA.
AC   Q9S495;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
```

```
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=20327582; PubMed=10869080;
RA   Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA   Russell P.W., Orndorff P.E.;
RT   "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT   mutants and identification of a novel binding phenotype.";
RL   J. Bacteriol. 182:4012-4021(2000).
DR   EMBL; AF154927; AAD44321.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT      65     65       Y -> C.
SQ   SEQUENCE   300 AA;  31533 MW;  48FFB2795A3CB23D CRC64;

  Query Match             97.1%;  Score 823;  DB 2;  Length 300;
  Best Local Similarity   97.5%;  Pred. No. 4.4e-70;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFYHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 9
Q9F6Z7
ID   Q9F6Z7      PRELIMINARY;      PRT;   300 AA.
AC   Q9F6Z7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Type 1 fimbrial adhesin subunit FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LF82;
RX   MEDLINE=21150472; PubMed=11251843;
RA   Boudeau J., Barnich N., Darfeuille-Michaud A.;
RT   "Type 1 pili-mediated adherence of Escherichia coli strain LF82
RT   isolated from Crohn's disease is involved in bacterial invasion of
RT   intestinal epithelial cells.";
RL   Mol. Microbiol. 39:1272-1284(2001).
DR   EMBL; AF288194; AAG24827.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   300 AA;  31445 MW;  BA040E9DD2137E43 CRC64;

  Query Match             96.9%;  Score 822;  DB 2;  Length 300;
  Best Local Similarity   96.3%;  Pred. No. 5.5e-70;
  Matches  155;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 10
O87634
ID   O87634      PRELIMINARY;      PRT;   300 AA.
AC   O87634;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IHE3034;
RA   Pouttu R., Korhonen T.K.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF089840; AAC35864.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   300 AA;  31429 MW;  D789457A5F5255AE CRC64;

  Query Match             96.6%;  Score 819;  DB 2;  Length 300;
  Best Local Similarity   95.7%;  Pred. No. 1.1e-69;
  Matches  154;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||:|||||||:|||||||:|||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 11
Q9AP05
ID   Q9AP05      PRELIMINARY;      PRT;   303 AA.
AC   Q9AP05;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RS218;
RA   Elliott S.J., Kim K.S.;
RT   "Virulence factors of meningitic Escherichia coli.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF317710; AAG60686.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   303 AA;  31773 MW;  4F615EA5D5E6763C CRC64;

  Query Match             96.6%;  Score 819;  DB 2;  Length 303;
```

```
  Best Local Similarity   95.7%;  Pred. No. 1.1e-69;
  Matches  154;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||:|||||||:|||||||:|||||||
Db      50 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 109

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     110 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 169

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||| ||||||||||
Db     170 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 210


RESULT 12
Q9S6R1
ID   Q9S6R1      PRELIMINARY;      PRT;   300 AA.
AC   Q9S6R1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Type 1 fimbriae adhesin, precursor polypeptide.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=88038337; PubMed=2890081;
RA   Klemm P., Christiansen G.;
RT   "Three fim genes required for the regulation of length and mediation
RT   of adhesion of Escherichia coli type 1 fimbriae.";
RL   Mol. Gen. Genet. 208:439-445(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=96235989; PubMed=8636962;
RA   Marc D., Dho-Moulin M.;
RT   "Analysis of the fim cluster of an avian O2 strain of Escherichia
RT   coli: serogroup-specific sites within fimA and nucleotide sequence of
RT   fimI.";
RL   J. Med. Microbiol. 44:444-452(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=96125229; PubMed=8550444;
RA   Klemm P., Tong S., Nielsen H., Conway T.;
RT   "The gntP gene of Escherichia coli involved in gluconate uptake.";
RL   J. Bacteriol. 178:61-67(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G.III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=98439336; PubMed=9766199;
RA   Marc D., Arne P., Bree A., Dho-Moulin M.;
RT   "Colonization ability and pathogenic properties of a fim- mutant of an
RT   avian strain of Escherichia coli.";
RL   Res. Microbiol. 149:473-485(1998).
DR   EMBL; AJ225176; CAA12423.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   300 AA;  31459 MW;  EA040CEDD2137FE8 CRC64;

  Query Match           96.2%;  Score 816;  DB 2;  Length 300;
  Best Local Similarity   95.7%;  Pred. No. 2e-69;
  Matches  154;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||| ||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      47 PAVNVGQKLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 13
Q08858
ID   Q08858      PRELIMINARY;      PRT;   243 AA.
AC   Q08858;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   FIMH protein precursor.
GN   FIMH.
OS   Klebsiella pneumoniae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
OX   NCBI_TaxID=573;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89155420; PubMed=2563996;
RA   Gerlach G.F., Clegg S., Allen B.L.;
RT   "Identification and characterization of the genes encoding the type 3
RT   and type 1 fimbrial adhesins of Klebsiella pneumoniae.";
RL   J. Bacteriol. 171:1262-1270(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Nichols N.N., Clegg S.;
RL   Submitted (AUG-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
CC       ADHESION OF TYPE 1 FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION
CC       OF FIMBRIAE).
DR   EMBL; L23111; AAA25063.1; -.
KW   Fimbria; Signal.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    243       FIMH PROTEIN.
SQ   SEQUENCE   243 AA;  25977 MW;  C8394D7C44A98AE7 CRC64;

  Query Match           89.2%;  Score 756;  DB 2;  Length 243;
  Best Local Similarity   86.3%;  Pred. No. 7.7e-64;
  Matches  139;  Conservative   12;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||:|||||||:|:|||||
Db      49 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGTSYPFP 108

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||:|| ||:|:|||||||  |||||||:|||: | |||||||||| |||||||| |||:
Db     109 TTTETARVIYDSRTDKPWPAVLYLTPVSTAGGVAITAGSLIAVLILHQTNNYNSDSFQFI 168

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||| ||: :||||:
Db     169 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSMAVPLTVH 209


RESULT 14
Q9F5R9
ID   Q9F5R9      PRELIMINARY;      PRT;   166 AA.
```

```
AC   Q9F5R9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RA   Harris S.L., Spears P.A., Havell E.A., Hamrick T.S., Horton J.R.,
RA   Orndorff P.E.;
RT   "Isolation and characterization of Escherichia coli type 1 pili
RT   mutants that have altered binding specificities.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF306537; AAG30927.1; -.
SQ   SEQUENCE   166 AA;  17826 MW;  2E64EEA80AE286DC CRC64;

  Query Match             72.9%;  Score 618;  DB 2;  Length 166;
  Best Local Similarity   98.3%;  Pred. No. 5.9e-51;
  Matches  118;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166


RESULT 15
Q8XAX2
ID   Q8XAX2      PRELIMINARY;      PRT;   304 AA.
AC   Q8XAX2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative adhesin, similar to FimH protein.
GN   Z2206 OR ECS2107.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005354; AAG56266.1; -.
DR   EMBL; AP002557; BAB35530.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
KW   Complete proteome.
SQ   SEQUENCE   304 AA;  32041 MW;  EED538023D95AFD5 CRC64;

  Query Match             45.3%;  Score 384;  DB 16;  Length 304;
  Best Local Similarity   47.9%;  Pred. No. 2e-28;
  Matches   78;  Conservative   29;  Mismatches   54;  Indels    2;  Gaps    2;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETI-TDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPF 59
           ||:  |||||||||  | | |||     ||:: | :|||: | | :: |:: ::  :|||
Db      48 PVIQPGQNLVVDLSQHISCWNDYGGWYDTDHINLVQGSAFAGSLQSYKGSLYWNNVTYPF 107

Qy      60 PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNS-DDFQ 118
           | |: |  :    :|  | |: ||:||| :|||:||||| :|| : : :     | :
Db     108 PLTTNTNVLDIGDKTPMPLPLKLYITPVGAAGGVVIKAGEVIARIHMYKIATLGSGNPRN 167

Qy     119 FVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           | ||| :|| ||:|||| | :|:||| |||: ||  ||| ||
Db     168 FTWNIISNNSVVMPTGGCTVDSRNVTVNLPDFPGSAEIPLGVY 210


RESULT 16
Q9XBV8
ID   Q9XBV8      PRELIMINARY;      PRT;   129 AA.
AC   Q9XBV8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=20327582; PubMed=10869080;
RA   Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA   Russell P.W., Orndorff P.E.;
RT   "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT   mutants and identification of a novel binding phenotype.";
RL   J. Bacteriol. 182:4012-4021(2000).
DR   EMBL; AF154929; AAD44323.1; -.
SQ   SEQUENCE   129 AA;  14142 MW;  BDE1D4D0E11EC536 CRC64;

  Query Match             30.9%;  Score 262;  DB 2;  Length 129;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-17;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTV 50
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTV 96


RESULT 17
Q53298
ID   Q53298      PRELIMINARY;      PRT;   299 AA.
AC   Q53298;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   S fimbrial adhesin minor subunit SfaH.
GN   SFAH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93138776; PubMed=8093693;
RA   Hacker J., Kestler H., Hoschutzky H., Jann K., Lottspeich F.,
RA   Korhonen T.K.;
```

```
RT   "Cloning and characterization of the S fimbrial adhesin II complex of
RT   an Escherichia coli O18:K1 meningitis isolate.";
RL   Infect. Immun. 61:544-550(1993).
DR   EMBL; S53210; AAB25047.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   299 AA;  31972 MW;  8EC6E8E84E9067C7 CRC64;

  Query Match             24.4%;  Score 207;  DB 2;  Length 299;
  Best Local Similarity   32.6%;  Pred. No. 1.2e-11;
  Matches   56;  Conservative   22;  Mismatches   82;  Indels   12;  Gaps    5;

Qy       1 PVVNVGQNL-VVDLSTQIFCHN-DYPETITDYVTLQRGSAYGGVLS-------NFSGTVK 51
             |||   ::: |:|||  : | | |      ||: : :|| :   |        :|:
Db      34 PVVQYDKSISVLDLSQLVSCQNEDSTGQNYDYLKILKGSGFSPALDTKTYGRLDFTSRPT 93

Qy      52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQ--T 109
               |     :     |    ||:|  ||| |     | ||  | |:| | : :  |
Db      94 GYARQLPLQFDLQVTEAFYQYGVWKPFPAKLYLYPAPGVFGKVINNGDLLATLYVNKFST 153

Qy     110 NNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGS-VPIPLTV 160
               : :  | |  || ||| : || | ||: :| | || | |  | :||||
Db     154 KGQEAGERNFTWRFYATNDVHIQTGTCRVSSNNVKVDLPSYPGGPVTVPLTV 205


RESULT 18
Q46686
ID   Q46686      PRELIMINARY;      PRT;   299 AA.
AC   Q46686; Q53398;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Minor F1652 fimbrial subunit H.
GN   F1652H OR FOCH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=4787 O115:V165:F165;
RX   MEDLINE=95202083; PubMed=7894716;
RA   Harel J., Jacques M., Fairbrother J.M., Bosse M., Forget C.;
RT   "Cloning of determinants encoding F165(2) fimbriae from porcine
RT   septicaemic Escherichia coli confirms their identity as F1C
RT   fimbriae.";
RL   Microbiology 141:221-228(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92073661; PubMed=1683712;
RA   van Die I., Kramer C., Hacker J., Bergmans H., Jongen W., Hoekstra W.;
RT   "Nucleotide sequence of the genes coding for minor fimbrial subunits
RT   of the F1C fimbriae of Escherichia coli.";
RL   Res. Microbiol. 142:653-658(1991).
DR   EMBL; U09804; AAA74946.1; -.
DR   EMBL; S68237; AAB20439.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   299 AA;  32155 MW;  259C745174C59D05 CRC64;

  Query Match             23.5%;  Score 199.5;  DB 2;  Length 299;
  Best Local Similarity   33.3%;  Pred. No. 6.1e-11;
  Matches   54;  Conservative   24;  Mismatches   73;  Indels   11;  Gaps    6;

Qy      10 VVDLSTQIFCHN-DYPETITDYVTLQRGSAYGGVL-----SNFSGTVKYSGSSYPFPTTS 63
           |:||:  : | | |      ||: :::|: :   |          | : || |   |
Db      44 VLDLNQLVLCQNEDASGQNYDYLRVRQGTGFSPSLDAKTYGRLDFTNRLSGYSQTLPLQQ 103

Qy      64 ET-PRVVY-NSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQ--TNNYNSDDFQF 119
           :| |   |      ||:|  :|| |     | :| || |:| : : :  |    : : |
Db     104 DTKPTEAYWQYGVWKPFPAKMYLYPEPGVFGKLIHAGELVATVYVNKFSTMGQEAGERNF 163

Qy     120 VWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGS-VPIPLTV 160
               |  || ||| : || | ||: :| | || | |  | :||||
Db     164 TWRFYATNDVYIQTGTCRVSSNNVKVDLPSYPGGPVTVPLTV 205


RESULT 19
Q9L1F8
ID   Q9L1F8      PRELIMINARY;      PRT;   684 AA.
AC   Q9L1F8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative neutral zinc metalloprotease.
GN   SCO5447 OR SC3D11.04C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Brown S.P., Harris D.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL157916; CAB76001.1; -.
DR   HSSP; P05806; 1NPC.
DR   InterPro; IPR001570; Peptidase_M4.
DR   InterPro; IPR005075; Pep_M4_propep.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01447; Peptidase_M4; 1.
DR   Pfam; PF02868; Peptidase_M4_C; 1.
DR   Pfam; PF03413; Pep_M4_propep; 1.
DR   PRINTS; PR00730; THERMOLYSIN.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Metalloprotease; Protease.
SQ   SEQUENCE   684 AA;  71289 MW;  594F70DDAB732BCF CRC64;

  Query Match             10.5%;  Score 89;  DB 16;  Length 684;
  Best Local Similarity   27.6%;  Pred. No. 5.2;
  Matches   40;  Conservative   17;  Mismatches   64;  Indels   24;  Gaps    4;

Qy      37 SAYGGV--------------LSNFSGTVKYSGSSYPFPTTSETPRV-VYNSRTDKPWPV 80
           |  |||               || |||   :| :|  ||:   |   :  :: :| |
Db     453 SGIGGVDVHYSSGPANHFFYLLSEGSGTKTINGVTYNSPTSDGLPVTGIGRAKAEKIWFR 512

Qy      81 ALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCD--- 137
```

```
          ||     |:      : |:| |   |  |   :| ::  | | :|  :|    || |
Db    513 ALTTKFTSTTNYAGARTGTLAAAGELYGT---DSAEYTAVANAWAGINVGARPGGGDPGD 569

Qy    138 --VSARDVTVTLPDYRGSVPIPLTV 160
                | :|||  :|   |||
Db    570 GTTFESTTNVNIPDYGSAVTSSLTV 594


RESULT 20
Q9YN73
ID   Q9YN73      PRELIMINARY;      PRT;   315 AA.
AC   Q9YN73;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 34.8 kDa protein.
OS   Vaccinia virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10245;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LIVP-2;
RA   Babkin I.V., Petrov N.A., Shchelkunov S.N.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99045; CAB11743.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   315 AA;  34836 MW;  97F613E87D577708 CRC64;

  Query Match             10.4%;  Score 88.5;  DB 12;  Length 315;
  Best Local Similarity   22.7%;  Pred. No. 2.2;
  Matches   42;  Conservative   21;  Mismatches   49;  Indels   73;  Gaps    9;

Qy      2 VVNVGQNLVVDLSTQIFCHN--------DY--------------PETITDYVTLQRGSAY 39
          :||      :|: .     |:        ||              || ||| |
Db    127 IVNTDSESTIDIMLSGSTHSPETSSEKPDYIDNSNCSSVFEIATPEPITDNV-------- 178

Qy     40 GGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
           : : || |          || |    || |||  : :   |      :  | ||
Db    179 ----EDHTDTVTYTSDSINTVSATSGES----TTDETPEPITDKEEDHTVTDTVSYTTVS 230

Qy     89 SAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPT--GGCDVSARDVTVT 146
          :: |:|           |    :||   ::: | :|| | ||  ||        | :
Db    231 TSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPPTTVGGS-------TTS 268

Qy    147 LPDYR 151
          : :|:
Db    269 ISNYK 273


RESULT 21
Q9YN68
ID   Q9YN68      PRELIMINARY;      PRT;   270 AA.
AC   Q9YN68;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   Hypothetical 29.8 kDa protein.
OS   Cowpox virus (CPV).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10243;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HAMBURG-1985;
RA   Babkin I.V., Petrov N.A., Shchelkunov S.N.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99050; CAB11748.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   270 AA;  29822 MW;  2B952EC06A8445E6 CRC64;

  Query Match             10.3%;  Score 87;  DB 12;  Length 270;
  Best Local Similarity   22.7%;  Pred. No. 2.5;
  Matches   41;  Conservative   22;  Mismatches   52;  Indels   66;  Gaps    9;

Qy      2 VVNVGQNLVVDLSTQIFCHN--------DY--------------PETITDYVTLQRGSAY 39
          :||      :|:      |:        ||              || ||| |
Db     83 IVNTDSESTIDIILSGSTHSPETSSEKPDYIDNSNCSSVFEIATPEPITDNV-------- 134

Qy     40 GGVLSNFSGTVKYSGSSYPF-------PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           : : || |:  |          ||| |||  :  : |     :  | ||:: |
Db    135 ----EDHTDTVTYTSDSINTVSATSGEPTTDETPEPI-TDKEDHTVTDTVSYTTVSTSSG 189

Qy     93 LVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPT--GGCDVSARDVTVTLPDY 150
          :|              |    :||   ::: | :|| | ||  ||        | :: :|
Db    190 IV-------------TTKSTTDDAD-LYDTYNDNDTVPPTTVGGS-------TTSISNY 227

Qy    151 R 151
          :
Db    228 K 228


RESULT 22
Q9YN66
ID   Q9YN66      PRELIMINARY;      PRT;   313 AA.
AC   Q9YN66;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 34.4 kDa protein.
OS   Monkeypox virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10244;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZAIRE 77-0666;
RA   Babkin I.V., Petrov N.A., Shchelkunov S.N.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99052; CAB11750.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   313 AA;  34434 MW;  329993BF534B33CC CRC64;

  Query Match             10.1%;  Score 85.5;  DB 12;  Length 313;
  Best Local Similarity   25.5%;  Pred. No. 4.2;
  Matches   36;  Conservative   17;  Mismatches   37;  Indels   51;  Gaps    7;

Qy     24 PETITDYVTLQRGSAYGGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNS 72
          || ||| |             : : || |          || |    || |||  : :
Db    169 PEPITDNV-----------EDHTDTVTYTSDSINTVNASSGES----TTDETPEPITDK 212

Qy     73 RTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVP 132
            |      :  | ||:: |:|             |    :||   ::: | :|| | |
Db    213 EEDHTVTDTVSYTTVSTSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPP 257

Qy    133 T--GGCDVSARDVTVTLPDYR 151
          |  ||        | :: :|:
Db    258 TTVGGS-------TTSISNYK 271


RESULT 23
Q8V4S6
ID   Q8V4S6      PRELIMINARY;      PRT;   313 AA.
AC   Q8V4S6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   B2R.
```

```
GN   B2R.
OS   Monkeypox virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10244;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZAIRE-96-I-16;
RX   MEDLINE=21592287; PubMed=11734207;
RA   Shchelkunov S.N., Totmenin A.V., Babkin I.V., Safronov P.F.,
RA   Ryazankina O.I., Petrov N.A., Gutorov V.V., Uvarova E.A.,
RA   Mikheev M.V., Sisler J.R., Esposito J.J., Jahrling P.B., Moss B.,
RA   Sandakhchiev L.S.;
RT   "Human monkeypox and smallpox viruses: genomic comparison.";
RL   FEBS Lett. 509:66-70(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ZAIRE-96-I-16;
RA   Shchelkunov S.N., Totmenin A.V., Safronov P.F., Gutorov V.V.,
RA   Ryazankina O.I., Petrov N.A., Babkin I.V., Uvarova E.A.,
RA   Esposito J.J., Moss B., Sisler J.R., Jahrling P.B., Sandakhchiev L.S.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF380138; AAL40621.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
SQ   SEQUENCE   313 AA;  34452 MW;  A7D916067F56ACF0 CRC64;

  Query Match             10.1%;  Score 85.5;  DB 12;  Length 313;
  Best Local Similarity   25.5%;  Pred. No. 4.2;
  Matches   36;  Conservative   17;  Mismatches   37;  Indels   51;  Gaps    7;

Qy      24 PETITDYVTLQRGSAYGGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNS 72
           || ||| |             : : || |          || |    || |||  : :
Db     169 PEPITDNV------------EDHTDTVTYTSDSINTVNASSGES----TTDETPEPITDK 212

Qy      73 RTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVP 132
             |      :  | ||:: |:|              |    :||   ::: | :|| | |
Db     213 EEDHTVTDTVSYTTVSTSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPP 257

Qy     133 T--GGCDVSARDVTVTLPDYR 151
           |  ||        | :: :|:
Db     258 TTVGGS-------TTSISNYK 271


RESULT 24
O26812
ID   O26812      PRELIMINARY;      PRT;  1755 AA.
AC   O26812;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cell surface glycoprotein (S-layer protein).
GN   MTH716.
OS   Methanobacterium thermoautotrophicum.
OC   Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC   Methanobacteriaceae; Methanothermobacter.
OX   NCBI_TaxID=187420;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DELTA H;
RX   MEDLINE=98037514; PubMed=9371463;
RA   Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA   Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA   Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA   Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA   Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA   McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA   Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT   "Complete genome sequence of Methanobacterium thermoautotrophicum
RT   deltaH: functional analysis and comparative genomics.";
RL   J. Bacteriol. 179:7135-7155(1997).
DR   EMBL; AE000851; AAB85221.1; -.
DR   InterPro; IPR000601; PKD_domain.
DR   Pfam; PF00801; PKD; 2.
DR   SMART; SM00089; PKD; 2.
DR   PROSITE; PS50093; PKD; 2.
KW   Complete proteome.
SQ   SEQUENCE   1755 AA;  191961 MW;  02C446C99BB6F3DA CRC64;

  Query Match             9.9%;  Score 84;  DB 17;  Length 1755;
  Best Local Similarity   23.9%;  Pred. No. 51;
  Matches   44;  Conservative   25;  Mismatches   61;  Indels   54;  Gaps    9;

Qy      26 TITDYVTLQRGSAY-------------------------GGVLSNFSGTVKYSGSSYP 58
             |  :|:|: :|  :                         | |  |::  :| :| :
Db    1423 TKENYITVYQGPQFTLENLTVIPTSGVAPLNITVSANLTNSGDVAGNYTAQLKVNGQTVA 1482

Qy      59 FPTTS---------ETPRV----VYNSRTDKPWPVAL-YLTPVS-SAGGLVIKAGSLIAV 103
             | :         |  ||    |||   |   |||:  | | : : | | :   | :|
Db    1483 TKTIAVPAGQTVSVEFSRVLTAGVYNVTIDDLAPVAVTVLKPANITVGNLTVTPKSGVAP 1542

Qy     104 LILRQTNNY-NSDDF------QFVWN--IYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           | :  | |  |: |          ||  |    |   : || ||  | : : |:|     |:
Db    1543 LNITVTANLRNTGDLAGNYTAQFKVNGQTVATKTIAVPAGGVGVVSFNYTLT---SAGNY 1599

Qy     155 PIPL 158
           |: :
Db    1600 PVAI 1603


RESULT 25
Q05352
ID   Q05352      PRELIMINARY;      PRT;  1114 AA.
AC   Q05352;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Immuno-dominant variable surface antigen.
OS   Entamoeba histolytica.
OC   Eukaryota; Entamoebidae; Entamoeba.
OX   NCBI_TaxID=5759;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90354789; PubMed=1696956;
RA   Edman U., Meraz M.A., Rausser S., Agabian N., Meza I.;
RT   "Characterization of an immuno-dominant variable surface antigen from
RT   pathogenic and nonpathogenic Entamoeba histolytica.";
RL   J. Exp. Med. 172:879-888(1990).
DR   EMBL; X55028; CAA38847.1; -.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1114 AA;  125570 MW;  0445F99609313E1D CRC64;

  Query Match             9.8%;  Score 83.5;  DB 5;  Length 1114;
  Best Local Similarity   23.6%;  Pred. No. 32;
  Matches   38;  Conservative   32;  Mismatches   58;  Indels   33;  Gaps    7;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSY-PF 59
           |: ::| :| :: :|| :  ||          | : || :||::   | |  :: | | |
Db     214 PIDSIGFDLGLN-TTQPYIIND---------TFKIGSPFGGMIYLRSDTT-FTNSFYVTF 262

Qy      60 PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQF 119
             | : ||  |:: |   |      :| | | :  :     |:|   |  : :| |:
Db     263 SNVGRAPIINYNITTNEEWNSVL-----RNAPGNVAEIRTPGNRLVLTSRNIRSLEDAQY 317

Qy     120 VWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTV 160
           : : :             :|  :  ||| :      ||:|:
Db     318 ISDFWLK----------AISISNYAVTLEN------IPITL 342


RESULT 26
Q9JF47
ID   Q9JF47      PRELIMINARY;      PRT;   315 AA.
AC   Q9JF47;
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TA66R.
OS   Vaccinia virus (strain Tian Tan).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10253;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIAN TAN;
RA   Jin Q., Hou Y.D., Cheng N.H., Yao E.M., Cheng S.X., Yang X.K.,
RA   Jing D.Y., Yu W.H., Yuan J.S., Ma X.J.;
RT   "Complete genomic sequence of vaccinia virus (Tian Tan strain).";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF095689; AAF34065.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
SQ   SEQUENCE   315 AA;  34758 MW;  E04944814085CAA4 CRC64;

  Query Match             9.7%;  Score 82.5;  DB 12;  Length 315;
  Best Local Similarity   23.8%;  Pred. No. 8.2;
  Matches   38;  Conservative   22;  Mismatches   43;  Indels   57;  Gaps    8;

Qy       4 NVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKY---------- 52
           |:  : | :::|        || ||| |           : : || |
Db     159 NLNCSSVFEIAT--------PEPITDNV-----------EDHTDTVTYTSDSINTVSAS 198

Qy      53 SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNY 112
           || |    || |||  : :  |      :  | ||:: |:|              |
Db     199 SGES----TTDETPEPITDKEEDHTVTDTVSYTTVSTSSGIV-------------TTKS 240

Qy     113 NSDDFQFVWNIYANNDVVVPTG-GCDVSARDVTVTLPDYR 151
           :||   ::: | :|| |  |  ||       | :: :|:
Db     241 TTDDAD-LYDTYNDNDTVPSTTVGCS------TTSISNYK 273


RESULT 27
O58048
ID   O58048      PRELIMINARY;      PRT;  1155 AA.
AC   O58048;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   1155AA long hypothetical PYROLYSIN.
GN   PH0310.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=53953;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE=98344137; PubMed=9679194;
RA   Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA   Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA   Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA   Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA   Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA   Masuchi Y., Shizuya H., Kikuchi H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
DR   EMBL; AP000001; BAA29383.1; -.
DR   InterPro; IPR000668; Peptidase_C1.
DR   InterPro; IPR000169; SHprot_acsite.
DR   Pfam; PF00112; Peptidase_C1; 1.
DR   ProDom; PD000158; Peptidase_C1; 1.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   1155 AA;  130098 MW;  092F4EC73144AF79 CRC64;

  Query Match             9.7%;  Score 82.5;  DB 17;  Length 1155;
  Best Local Similarity   24.5%;  Pred. No. 42;
  Matches   46;  Conservative   24;  Mismatches   59;  Indels   59;  Gaps   11;

Qy      14 STQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPR------ 67
           :| :: || | :   ||::      |  :||| |    |   ::|||   |:|
Db     566 NTIVWFHN-YLDFWIDYIS---DKEYNAILSNLSLLESYLPQAHPFP---ESPMAFDVSQ 618

Qy      68 ----VVY--NSRTDKPWPVALYLTP----------------VSSAGGLVIKAGSLIAVL 104
               :||    :   |  |  |: |                 : | ||     |||
Db     619 LSDYLVYYMEDKNSTPTEVTFYVQPKDILPDNFTGELYNFTLLISIGGGEYKVLGSL---- 674

Qy     105 ILRQTNNYNSDDFQFV----------WNIYANNDVVVPTGGCDVSARDVTVTLPDYR--- 151
             | :  | :|          ::: | |:| |  |  :||   : : | | |
Db     675 ----EGNVSIPDGGYVTVPLEIPVVSYDV-APNNVTVQYGNFNVSIYSI-IPLEDARIVI 728

Qy     152 GSVPIPLT 159
           |:   |||
Db     729 GNESYPLT 736


RESULT 28
Q9X6L5
ID   Q9X6L5      PRELIMINARY;      PRT;  383 AA.
AC   Q9X6L5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 40.7 kDa protein.
OS   Klebsiella oxytoca.
OG   Plasmid pACM1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
OX   NCBI_TaxID=571;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20334746; PubMed=10873523;
RA   Preston K.E., Radomski C.C.A., Venezia R.A.;
RT   "Nucleotide sequence of a 7-kb fragment of pACM1 encoding an IncM DNA
RT   primase and other putative proteins associated with conjugation.";
RL   Plasmid 44:12-23(2000).
DR   EMBL; AF139719; AAD33810.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   383 AA;  40703 MW;  5F338A503B65911A CRC64;

  Query Match             9.7%;  Score 82;  DB 2;  Length 383;
  Best Local Similarity   29.4%;  Pred. No. 12;
  Matches   35;  Conservative   18;  Mismatches   54;  Indels   12;  Gaps    6;

Qy      47 SGTVKYS-GSSYP----FPTTSETPRVVYNSRTDKPWPVA---LYLTPVSSAGGLVIKAG 98
           | ||  | |:| |    ||  : |  | ::  |  || :          :| :| | | :
Db     163 SQTVSLSAGASIPQVRVFPNQATT--VTFSDATGHPWVLGAPPYNSSPCASGGELCVGYI 220

Qy      99 SLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIP 157
             ||  :: || | | :   :    |   |::   | : | :  || : |||  : ||
Db     221 PGSAVFTIQPTNAYASGNITVLLKGLA-TPVIINVKGAEPSVKSKTVDV-DYRLDLRIP 277


RESULT 29
Q89120
ID   Q89120      PRELIMINARY;      PRT;  315 AA.
AC   Q89120;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hemagglutinin precursor.
OS   Vaccinia virus (strain Tian Tan).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10253;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=TIAN TAN;
RA   Li Z., Jin Q., Yu W., Jin T., Hou Y.;
RT   "Complete nucleotide sequence of the hemagglutinin gene of tian tan
RT   strain of vaccinia virus.";
RL   Ping Tu Hsueh Pao 5:1-9(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TIAN TAN;
RA   Xu L.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U25662; AAA74188.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
KW   Signal.
FT   SIGNAL        1     16       POTENTIAL.
SQ   SEQUENCE   315 AA;  34772 MW;  794DC90B57879D53 CRC64;

  Query Match             9.6%;  Score 81.5;  DB 12;  Length 315;
  Best Local Similarity   23.8%;  Pred. No. 10;
  Matches   38;  Conservative   22;  Mismatches   43;  Indels   57;  Gaps    8;

Qy       4 NVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKY---------- 52
           |:  : | :::|        || ||| |            : : || |
Db     159 NLNCSSVFEIAT--------PEPITDNV-----------EDHTDTVTYTSDSINTVSAT 198

Qy      53 SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNY 112
           || |    || |||  : :   |      :  | ||:: |:|              |
Db     199 SGES----TTDETPEPITDKEEDHTVTDTVSYTTVSTSSGIV-------------TTKS 240

Qy     113 NSDDFQFVWNIYANNDVVVPTG-GCDVSARDVTVTLPDYR 151
            :||   ::: | :|| |  |  ||       | :: :|:
Db     241 TTDDGD-LYDTYNDNDTVPSTTVGCS------TTSISNYK 273


RESULT 30
Q82496
ID   Q82496      PRELIMINARY;      PRT;   566 AA.
AC   Q82496;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hemagglutinin.
GN   HA.
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=11320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/SWINE/ANGE-GARDIEN/150/90(H3N2);
RX   MEDLINE=95205091; PubMed=7897358;
RA   Bikour M.H., Frost E.H., Deslandes S., Talbot B., Weber J.M.,
RA   Elazhary Y.;
RT   "Recent H3N2 swine influenza virus with haemagglutinin and
RT   nucleoprotein genes similar to 1975 human strains.";
RL   J. Gen. Virol. 76:697-703(1995).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; U07146; AAA85781.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63362 MW;  1FB8319A567E2FFF CRC64;

  Query Match             9.6%;  Score 81.5;  DB 12;  Length 566;
  Best Local Similarity   26.4%;  Pred. No. 21;
  Matches   39;  Conservative   17;  Mismatches   57;  Indels   35;  Gaps    7;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |       |::|  :  |||         |   : ||
Db     112 NCYPYDVPDYASLRSLVASSGTLEFINEGFNWTGVTQNGGSS----ACKRGPDNGFFSRL 167

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |   ||            |:||   |  :   || |||   ::| ::         |
Db     168 N--W---LY-----------KSGSTYPVQNVTMPNNDNSDKL-YIWGVHHPSTDKEQTN 209

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || :    |:    || |
Db     210 LYVQASGKVTVSTKRSQQTVIPNVGSRP 237


RESULT 31
Q40713
ID   Q40713      PRELIMINARY;      PRT;   824 AA.
AC   Q40713;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Protein kinase.
GN   OSPK10.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhao Y., Feng X.-H., Bottino P.J., Kung S.-D.;
RT   "Molecular and biochemical characterization of a receptor
RT   Serine/Threonine Kinase from Rice.";
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; L27821; AAA33915.1; -.
DR   InterPro; IPR001480; B_lectin.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003014; PAN.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR004040; STY_pkinase.
DR   Pfam; PF01453; Agglutinin; 1.
DR   Pfam; PF00024; PAN; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00108; B_lectin; 1.
DR   SMART; SM00221; STYKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   824 AA;  89870 MW;  EEEE844A0BEEDDC8 CRC64;

  Query Match             9.6%;  Score 81;  DB 10;  Length 824;
  Best Local Similarity   23.4%;  Pred. No. 38;
  Matches   39;  Conservative   29;  Mismatches   57;  Indels   42;  Gaps    9;

Qy       7 QNLVV--DLSTQIFCHNDYPETITDYVTLQRGSAY--GGVLSNFSGTVK---------- 51
           |::|:  | |: ::    :|   ||  ||  |  :  |  | : | ||:
Db     145 QSVVLGKDASSPLWQSFSHP---TD--TLLSGQNFIEGMTLMSKSNTVQNMTYTLQIKSG 199

Qy      52 ----YSGSSYPFP--TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLI 105
                |:|   | |  :  :  |:: |   |: :|  |     |::     ::|||:: |:
Db     200 NMILYAGFETPQPYWSAQQDSRIIVNKNGDRIYPANL----SSASWSFYDQSGSLLSQLV 255

Qy     106 LRQTNN-------YNSDDFQFVWNIYANN-----DVVVPTGGCDVSA 140
           : | |          ||     : :   |      : ||   ||: |
Db     256 IAQENANATLSAVLGSDGLIAFYMLQGGNGKSKFSITVPADSCDMPA 302


RESULT 32
```

```
Q9SE00
ID   Q9SE00      PRELIMINARY;      PRT;   473 AA.
AC   Q9SE00;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Purple acid phosphatase precursor (EC 3.1.3.2).
GN   PAP1.
OS   Ipomoea batatas (Sweet potato) (Batate).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
OX   NCBI_TaxID=4120;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99441212; PubMed=10510276;
RA   Schenk G., Ge Y., Carrington L.E., Wynne C.J., Searle I.R.,
RA   Carroll B.J., Hamilton S., de Jersey J.;
RT   "Binuclear metal centers in plant purple acid phosphatases: Fe-Mn in
RT   sweet potato and Fe-Zn in soybean.";
RL   Arch. Biochem. Biophys. 370:183-189(1999).
CC   -!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O = AN
CC       ALCOHOL + PHOSPHATE.
DR   EMBL; AF200825; AAF19821.1; -.
DR   HSSP; P80366; 4KBP.
DR   InterPro; IPR004843; M-ppestrase.
DR   InterPro; IPR004844; S/T_phosphtse.
DR   Pfam; PF00149; Metallophos; 1.
KW   Hydrolase; Iron; Signal; Zinc.
FT   SIGNAL        1     38       POTENTIAL.
FT   CHAIN        39    473       PURPLE ACID PHOSPHATASE.
SQ   SEQUENCE   473 AA;  53815 MW;  BAE4B807DADD95A7 CRC64;

  Query Match             9.5%;  Score 80.5;  DB 10;  Length 473;
  Best Local Similarity   37.9%;  Pred. No. 21;
  Matches   22;  Conservative   12;  Mismatches   19;  Indels    5;  Gaps    2;

Qy      30 YVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPV 87
           :  ::| ||:  |||::|| |||     |  |||  :|   :|:: || : |   |:
Db     274 WYAIKRASAHIIVLSSYSGFVKYSPQYKWF--TSELEKV---NRSETPWLIVLVHAPL 326


RESULT 33
Q9XC46
ID   Q9XC46      PRELIMINARY;      PRT;  1304 AA.
AC   Q9XC46;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Outer membrane protein B (Fragment).
GN   OMPB.
OS   Rickettsia australis.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PHS;
RX   MEDLINE=20487299; PubMed=11034486;
RA   Stenos J., Walker D.H.;
RT   "The rickettsial outer-membrane protein A and B genes of Rickettsia
RT   australis, the most divergent rickettsia of the spotted fever group.";
RL   Int. J. Syst. Evol. Microbiol. 50:1775-1779(2000).
DR   EMBL; AF149109; AAD39532.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   InterPro; IPR001639; T2SP_C.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
DR   PRINTS; PR00810; BCTERIALGSPC.
DR   PRINTS; PR01656; VACCYTOTOXIN.
FT   NON_TER       1     1
SQ   SEQUENCE   1304 AA;  133283 MW;  ABFF9AAEEDA87701 CRC64;

  Query Match             9.5%;  Score 80.5;  DB 2;  Length 1304;
  Best Local Similarity   25.4%;  Pred. No. 75;
  Matches   33;  Conservative   13;  Mismatches   53;  Indels   31;  Gaps    4;

Qy      27 ITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTP 86
           | :  ||  |    |   |  |||  ||          ||  :|    :  :|    :|
Db     595 IVNATTLYAGI---GTTKNNQGTVTLSGG------IPNTPGTIYGLGVEHGYPKLKQVTF 645

Qy      87 VSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVT 146
           :    |    ||:||                    |:  |: | | |||   :  | |:|
Db     646 TTDYNNL----GSIIAT-----------------NVTINDGVTVTTGGVVGTDFDGTIT 683

Qy     147 LPDYRGSVPI 156
           |    |:  :
Db     684 LGSVNGNANV 693


RESULT 34
Q9KKB5
ID   Q9KKB5      PRELIMINARY;      PRT;  1620 AA.
AC   Q9KKB5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   OmpB (Fragment).
GN   OMPB.
OS   Rickettsia australis.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=787;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PHILLIPS;
RX   MEDLINE=20393643; PubMed=10939649;
RA   Roux V., Raoult D.;
RT   "Phylogenetic analysis of members of the genus Rickettsia using the
RT   gene coding the outer-membrane protein rOmpB (ompB).";
RL   Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
DR   EMBL; AF123709; AAF34112.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
FT   NON_TER    1620   1620
SQ   SEQUENCE   1620 AA;  164420 MW;  F9E9DB056A548A1C CRC64;

  Query Match             9.5%;  Score 80.5;  DB 2;  Length 1620;
  Best Local Similarity   25.4%;  Pred. No. 99;
  Matches   33;  Conservative   13;  Mismatches   53;  Indels   31;  Gaps    4;

Qy      27 ITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTP 86
           | :  ||  |    |   |  |||  ||          ||  :|    :  :|    :|
Db     918 IVNATTLYAGI---GTTKNNQGTVTLSGG------IPNTPGTIYGLGVEHGYPKLKQVTF 968

Qy      87 VSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVT 146
           :    |    ||:||                    |:  |: | | |||   :  | |:|
Db     969 TTDYNNL----GSIIAT-----------------NVTINDGVTVTTGGVVGTDFDGTIT 1006

Qy     147 LPDYRGSVPI 156
           |    |:  :
Db    1007 LGSVNGNANV 1016


RESULT 35
Q9YN72
ID   Q9YN72      PRELIMINARY;      PRT;   270 AA.
AC   Q9YN72;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   Hypothetical 29.8 kDa protein.
OS   Vaccinia virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10245;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=LIVP-1;
RA   Babkin I.V., Petrov N.A., Shchelkunov S.N.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99046; CAB11744.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   270 AA;  29798 MW;  EADC0F1B0567CCC2 CRC64;

  Query Match             9.4%;  Score 80;  DB 12;  Length 270;
  Best Local Similarity   22.7%;  Pred. No. 12;
  Matches   42;  Conservative   21;  Mismatches   48;  Indels   74;  Gaps   10;

Qy       2 VVNVGQNLVVDLSTQIFCHN--------DY--------------PETITDYVTLQRGSAY 39
           :||      :|:      |:       ||              || ||| |
Db      83 IVNTDSESTIDIILSGSTHSPETSSEKPDYIDNSNCSSVFEIATPEPITDNV-------- 134

Qy      40 GGVLSNFSGTVKY-----------SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
             : : || |          || |    || |||  :   : |      :  | ||
Db     135 ----EDHTDTVTYTSDSINTVSASSGES----TTDETPEPI-TDKEDHTVTDTVSYTTVS 185

Qy      89 SAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPT--GGCDVSARDVTVT 146
           :: |:|              |    :||   ::: | :|| | ||  ||        | :
Db     186 TSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPPTTVGGS-------TTS 223

Qy     147 LPDYR 151
           : :|:
Db     224 ISNYK 228


RESULT 36
Q96UK8
ID   Q96UK8      PRELIMINARY;      PRT;   519 AA.
AC   Q96UK8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Laccase 1 (EC 1.10.3.2).
GN   LAC1.
OS   Trametes versicolor (White-rot fungus).
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Aphyllophorales; Trametes.
OX   NCBI_TaxID=5325;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CU1;
RA   McClean K.H., O'Brien M.M., Dobson A.D.W.;
RT   "Trametes versicolor laccase (lac1) mRNA sequence.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY049725; AAL00887.1; -.
DR   InterPro; IPR001117; Cu-oxidase.
DR   Pfam; PF00394; Cu-oxidase; 3.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.
KW   Oxidoreductase.
SQ   SEQUENCE   519 AA;  55998 MW;  8C9C6934DEAE3E0A CRC64;

  Query Match             9.4%;  Score 80;  DB 3;  Length 519;
  Best Local Similarity   27.7%;  Pred. No. 27;
  Matches   38;  Conservative   21;  Mismatches   48;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db     249 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 304

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | :   | | ||        |:   |:|
Db     305 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGTRTPGGVDKA--------LKLAFNFNGT 355

Qy     116 DFQFVWNIYANNDVVVP 132
           :|      : ||    |
Db     356 NF------FINNASFTP 366


RESULT 37
Q89182
ID   Q89182      PRELIMINARY;      PRT;   315 AA.
AC   Q89182;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HAEMAGGLUTININ.
GN   A56R OR MVA165R.
OS   Vaccinia virus (strain Ankara).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=126794;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97080498; PubMed=8921843;
RA   Antoine G., Scheiflinger F., Holzer G., Langmann T., Falkner F.G.,
RA   Dorner F.;
RT   "Characterization of the vaccinia MVA hemagglutinin gene locus and its
RT   evaluation as an insertion site for foreign genes.";
RL   Gene 177:43-46(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANKARA;
RA   Antoine G., Scheiflinger F., Falkner F.G., Dorner F.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X91135; CAA62575.1; -.
DR   EMBL; U94848; AAB96543.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
SQ   SEQUENCE   315 AA;  34794 MW;  C7EE7D42D7E4E543 CRC64;

  Query Match             9.4%;  Score 79.5;  DB 12;  Length 315;
  Best Local Similarity   22.2%;  Pred. No. 16;
  Matches   41;  Conservative   21;  Mismatches   50;  Indels   73;  Gaps    9;

Qy       2 VVNVGQNLVVDLSTQIFCHN--------DY--------------PETITDYVTLQRGSAY 39
           :||      :|:      |:       ||              || ||| |
Db     127 IVNTDSESTIDIILSGSTHSPETSSEKPDYIDNSNCSSVFEIATPEPITDNV-------- 178

Qy      40 GGVLSNFSGTVKY-----------SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
             : : || |          || |    || |||  :   |      :  | ||
Db     179 ----EDHTDTVTYTSDSINTVSASSGES----TTDETPEPITDKEEDHTVTDTVSYTTVS 230

Qy      89 SAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPT--GGCDVSARDVTVT 146
           :: |:|              |    :||   ::: | :|| |  |  ||        | :
Db     231 TSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPSTTVGGS-------TTS 268

Qy     147 LPDYR 151
           : :|:
Db     269 ISNYK 273


RESULT 38
Q67126
ID   Q67126      PRELIMINARY;      PRT;   566 AA.
AC   Q67126;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hemagglutinin.
GN   HA.
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=11320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/SEAL/MA/3984/92;
RX   MEDLINE=95146951; PubMed=7844533;
RA   Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT   "The appearance of H3 influenza viruses in seals.";
```

```
RL   J. Gen. Virol. 76:199-203(1995).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; L32024; AAA64228.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63441 MW;  590576CB4CEE7D08 CRC64;

  Query Match             9.4%;  Score 79.5;  DB 12;  Length 566;
  Best Local Similarity   25.7%;  Pred. No. 33;
  Matches   38;  Conservative   17;  Mismatches   58;  Indels   35;  Gaps    7;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||     |    |   : ||
Db     112 NCYPYDVPDYASLRSLVASSGTLEFIAEGFTWTGVTQNGGSG----TCKRGPANGFFSRL 167

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY---ANND--- 128
           :  |              : |:||   || :   || | |   :|| ::    | :
Db     168 N--W--------------LTKSGSAYPVLNVTMPNNDNFDKL-YVWGVHHPSTNQEQTS 209

Qy     129 -VVVPTGGCDVSARDVTVTLPDYRGSVP 155
            |  :|   || |    |:   || |
Db     210 LYVQASGRVTVSTRRSQQTIIPNIGSSP 237


RESULT 39
Q9JFS2
ID   Q9JFS2      PRELIMINARY;      PRT;  1924 AA.
AC   Q9JFS2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   C15R.
GN   C15R.
OS   Ectromelia virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=12643;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MOSCOW;
RX   MEDLINE=20192152; PubMed=10725549;
RA   Chen N., Buller R.M.L., Wall E.M., Upton C.;
RT   "Analysis of host response modifier ORFs of ectromelia virus, the
RT   causative agent of mousepox.";
RL   Virus Res. 66:155-173(2000).
DR   EMBL; AF012825; AAC99574.1; -.
SQ   SEQUENCE   1924 AA;  216775 MW;  B34C50FEA260BB95 CRC64;

  Query Match             9.4%;  Score 79.5;  DB 12;  Length 1924;
  Best Local Similarity   18.1%;  Pred. No. 1.5e+02;
  Matches   46;  Conservative   36;  Mismatches   73;  Indels   99;  Gaps   10;

Qy       1 PVVNVGQNLVVDLSTQ----IFCHNDYPE-----------TITD--------------- 29
           || : |   :   : |    :|| ::| :           :|:|
Db    1286 PVGDYGTAKLYSATKQSRIAVFCTHNYDKRFKSDIIVLIFNSISDIRFSSIYTGDVNGRN 1345

Qy      30 --YVTLQRGSAYGGV-LSNFSGTVKYSGSSY-----------------PFPTTSETPR 67
             ::|| ||  |  :   |     |:|  :                   :  |  ||
Db    1346 RLFITLSRGMPYRSMYCDNRRPGCYYAGIPFNENSVESDLHYGPEIMLKETYDANSIDPR 1405

Qy      68 VVYNSRTDKPWPVALYLTPVSSAGG-------------------LVIKA---------- 97
           |:  |:|  | |:::    :   |                     : ||
Db    1406 VITKSKTHFPTPISVKFMVANLGNGYNKPENFWNDAKSKKRTYSAMTIKILPCTVRNKNV 1465

Qy      98 ------GSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGC----DVSARDVTVTL 147
                | :|: ::  |: :|:  |       |    |      |    |: |::||||
Db    1466 NFGYNYGHIISNMVYAQSTSYDYGD----GTNYTFKSVNRSDHECESILDLKAKEVTVTC 1521

Qy     148 PDYRGSVPIPLTVY 161
           | :  |:|  :: |
Db    1522 PAF--SIPRNISAY 1533


RESULT 40
O64302
ID   O64302      PRELIMINARY;      PRT;  421 AA.
AC   O64302;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   A-protein.
OS   bacteriophage M11.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Leviviridae;
OC   Allolevivirus.
OX   NCBI_TaxID=74336;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95239761; PubMed=7723040;
RA   Beekwilder M.J., Nieuwenhuizen R., van Duin J.;
RT   "Secondary structure model for the last two domains of single-stranded
RT   RNA phage Q beta.";
RL   J. Mol. Biol. 247:903-917(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96190948; PubMed=8609616;
RA   Beekwilder J., Nieuwenhuizen R., Poot R., van Duin J.;
RT   "Secondary structure model for the first three domains of Q beta RNA.
RT   Control of A-protein synthesis.";
RL   J. Mol. Biol. 256:8-19(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Beekwilder J., Nieuwenhuizen R., Poot R., van Duin J.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF052431; AAC06249.1; -.
SQ   SEQUENCE   421 AA;  48619 MW;  EF24627BF4EC7707 CRC64;

  Query Match             9.3%;  Score 79;  DB 9;  Length 421;
  Best Local Similarity   26.2%;  Pred. No. 25;
  Matches   34;  Conservative   15;  Mismatches   57;  Indels   24;  Gaps    4;

Qy      31 VTLQRGSAYGGVLSNFSGTVKY-SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSS 89
           :|:||   :| | :|  |    : :||  |     |      |     |  ||  ||| |
Db     267 ITVQRRHRWGIVYANREGYATFDNGSIRPVSDWKELANAFIN-----PGEVAWELTPYSF 321

Qy      90 AGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPD 149
                 |  | :|                 :    :| | |:|      |:  |  :|||
Db     322 VVDWFINVGDII---------------EQQKQLYQNIDIVDGYQRRDIRMR--SVTLKG 363

Qy     150 YRGSVPIPLT 159
             |  :|: :|
Db     364 VRNGIPVRVT 373


RESULT 41
Q9VFW7
ID   Q9VFW7      PRELIMINARY;      PRT;  985 AA.
AC   Q9VFW7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   CG8775 protein.
GN   CG8775.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003699; AAF54928.1; -.
DR   FlyBase; FBgn0038138; CG8775.
DR   InterPro; IPR001930; Ala_peptase.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01433; Peptidase_M1; 1.
DR   PRINTS; PR00756; ALADIPTASE.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   985 AA;  111664 MW;  56CAC0CF41172224 CRC64;

  Query Match           9.3%;  Score 79;  DB 5;  Length 985;
  Best Local Similarity   21.9%;  Pred. No. 73;
  Matches   43;  Conservative   24;  Mismatches   57;  Indels   72;  Gaps    9;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDY---VTLQRGSAYGGVLSNFSGTVKYSGSSY 57
           ||: :   |    | |    : | ||:|   :|  :|:|   :| | | | |   :
Db     502 PVLTIDATLA---SHPIVKSIESPAEITEYFDTITYSKGAALVRMLENLVGEEKLRNA-- 556

Qy      58 PFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGL--------------------VIK 96
             ||    | :|:: | :      ||| |    ||                    | |
Db     557 ---TTRYLVRHIYSTATTED-----YLTAVEEEEGLEFDVKQIMQTWTEQMGLPVVEVEK 608

Qy      97 AGS---LIAVLILRQTNNY----NSDDFQFVWN---------------IYANNDVVVPT 133
           :||    |    |   ::|     :  | : |:               |: :||
Db     609 SGSTYKLTQKRFLANEDDYAAEAEASSFNYRWSIPITYTSSINSEVQSLIFNHND----- 663

Qy     134 GGCDVSARDVTVTLPD 149
                 : |:|||:
Db     664 -------NEATITLPE 672


RESULT 42
Q9FX19
ID   Q9FX19      PRELIMINARY;      PRT;  1064 AA.
AC   Q9FX19;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative protein kinase.
GN   F12G12.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA   Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA   Buehler E., Chao Q., Chin C., Chiou J., Choi E., Gonzalez A.,
RA   Howng B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA   Lenz C., Liu A., Liu S., Mukharsky N., Pham P., Sakano H., Shinn P.,
RA   Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC015446; AAG12526.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR004838; NHtransf_1.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR004040; STY_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00560; LRR; 24.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00370; LRR; 23.
DR   SMART; SM00221; STYKc; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00105; AA_TRANSFER_CLASS_1; UNKNOWN_1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Transferase.
SQ   SEQUENCE   1064 AA;  115915 MW;  2E8FC24DFF4F053D CRC64;

  Query Match           9.3%;  Score 79;  DB 10;  Length 1064;
  Best Local Similarity   21.2%;  Pred. No. 81;
  Matches   45;  Conservative   21;  Mismatches   58;  Indels   88;  Gaps   10;

Qy      11 VDLSTQIFCHNDYPETITDYVTLQ-----RGSAYGGV----------------LSNFSGTV 50
           :|||   |  |  |||  :|   ||       | :| :               :|||| :
Db     602 LDLSYNTFTGN-IPETFSDLTQLQSLDLSSNSLHGDIKVLGSLTSLASLNISCNNFSGPI 660

Qy      51 KYSGSSYPFPTTSETPRVVYNSR------------------TDKPWPVALYLTPVSSAG 91
               | ||  |  |   : |:                        |  |||
Db     661 ----PSTPFFKTISTTSYLQNTNLCHSLDGITCSSHTGQNNGVKSPKIVAL--------- 707

Qy      92 GLVIKAGSLIAV-----LILRQTNNY----------NSDDFQFVWNIYANNDVVVPTGG 135
             || |   ||:     ||||  : |            ::|| : |          :|
Db     708 TAVILASITIAILAAWLLILRNNHLYKTSQNSSSSPSTAEDFSYPWTF-------IPFQK 760

Qy     136 CDVSARDVTVTLPD-----------YRGSVP 155
           ::  ::  :| |              |:  :|
Db     761 LGITVNNIVTSLTDENVIGKGCSGIVYKAEIP 792


RESULT 43
Q8T2A1
ID   Q8T2A1      PRELIMINARY;      PRT;  3763 AA.
AC   Q8T2A1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 416.9 kDa protein.
```

```
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115678; AAL92607.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   3763 AA;  416939 MW;  83E8D27C0164EDF6 CRC64;

  Query Match             9.3%;  Score 79;  DB 5;  Length 3763;
  Best Local Similarity   24.5%;  Pred. No. 3.9e+02;
  Matches   46;  Conservative   25;  Mismatches   75;  Indels   42;  Gaps    8;

Qy       5 VGQNLVVDLSTQIFCHN-----DYPETITDYVTLQRG-SAYGGVLSNFSGTVKYSGSS-- 56
Db    1853 IGQDTSVDLSNVIAGGSTVDGVDIPNVIAGGSTVDGGTSVNGGTGSTSTTTVSSSPSSSS 1912

Qy      57 ----------------YPFPTTSETPRVVYNSRTDKP-------WPVALYLTPV--SSAG 91
Db    1913 DIGSSSDISSEVSSSLSSSPSSSEQPSEQSSSSSEQPPEDSMEEYPVH-YYEPVFETTPG 1971

Qy      92 GLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYR 151
Db    1972 SYPLDSIKAERTDLTRNAAIVRAAPYAFGQKVF-----FIDNNGNEVGVPDVTVTLVNSN 2026

Qy     152 GSVPIPLT 159
Db    2027 G---FPVT 2031


RESULT 44
Q8U9U1
ID   Q8U9U1      PRELIMINARY;      PRT;   398 AA.
AC   Q8U9U1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Penicillin-binding protein dacf.
GN   DACF OR ATU3634 OR AGR_L_2387.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009292; AAL44446.1; ALT_INIT.
DR   EMBL; AE008320; AAK89763.1; -.
KW   Complete proteome.
SQ   SEQUENCE   398 AA;  42160 MW;  8155CAABE9565C5D CRC64;

  Query Match             9.3%;  Score 78.5;  DB 16;  Length 398;
  Best Local Similarity   24.2%;  Pred. No. 26;
  Matches   43;  Conservative   26;  Mismatches   64;  Indels   45;  Gaps    8;

Qy       1 PVVNVGQNLVVDLST-QIFCHND-----YPETITDYVTLQRGSAYGGVLSNFSGTVKYSG 54
Db      56 PVANANPKMVVDVKTGKVISHQEAFRKWYPASLTKLMT-----AYIAFSQMKAGKI---- 106

Qy      55 SSYPFPTTSETPRVVYNSRTDKPWPVALY-----LTPVSSAGGLVIKAGSLIAVLILRQT 109
Db     107 -------SPQTEVVMSKKAADQPASKMYFKPGQKLTMDSALKLLLIKSANDIAVAI-AET 158

Qy     110 NNYNSDDFQFVWNIYAN----------NDVVVPTGGCDVSARDVTV-------TLPDY 150
Db     159 IGGTSDNFVAQMNAQAQRLGMSSTHYVNPNGLPGKGQYTTARDLALLALIIKREFPEY 216


RESULT 45
Q93SH4
ID   Q93SH4      PRELIMINARY;      PRT;   782 AA.
AC   Q93SH4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   HmuR protein.
GN   HMUR.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Bradyrhizobium group; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=110SPC4;
RA   Nienaber A., Hennecke H., Fischer H.M.;
RT   "Discovery of a haem uptake system in the soil bacterium
RT   Bradyrhizobium japonicum.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ311165; CAC38746.1; -.
DR   InterPro; IPR000531; TonB_boxC.
DR   Pfam; PF00593; TonB_boxC; 1.
DR   PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
SQ   SEQUENCE   782 AA;  83713 MW;  F8FC1EBFE568AA87 CRC64;

  Query Match             9.3%;  Score 78.5;  DB 2;  Length 782;
  Best Local Similarity   25.0%;  Pred. No. 61;
  Matches   37;  Conservative   17;  Mismatches   65;  Indels   29;  Gaps    7;

Qy      21 NDYPETITDYVTLQRGSA-YGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWP 79
Db     313 NRGPTTTQALIALNRGSSVYASDAKNYSGTVTWN-YSLPSDNLFDWHMSVYGNRTDNDQT 371

Qy      80 VALY--LTP---------------VSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQ--FV 120
Db     372 KTYHYGTTPSAYCNGGFGNNVSGCVGDKRGYVLNTYGVDA----NNTTRFNVGDWRNALT 427

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLP 148
Db     428 WGVDAFQDDVITTD----SRGNSNITTP 451


Search completed: November 28, 2002, 17:21:30
Job time : 30 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 17:17:25 ; Search time 44 Seconds
(without alignments)
351.765 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	834	98.3	300	2	S56545	fimbrial protein f
2	824	97.2	300	2	G91288	hypothetical prote
3	824	97.2	300	2	B86130	hypothetical prote
4	756	89.2	302	2	A32801	fimbrial adhesin p
5	384	45.3	304	2	A64904	probable fimbrial
6	384	45.3	304	2	C90892	probable adhesin [
7	384	45.3	304	2	F85725	probable adhesin,
8	207	24.4	299	2	C49233	S fimbrial adhesin
9	205	24.2	291	2	S15927	sfaH protein precu
10	199.5	23.5	299	2	I76900	F1652 minor fimbri
11	86.5	10.2	550	1	HMIV98	hemagglutinin prec
12	85.5	10.1	315	1	HNVZVV	hemagglutinin prec
13	85.5	10.1	550	1	HMIV33	hemagglutinin prec
14	84	9.9	1797	2	F69195	cell surface glyco
15	83.5	9.8	567	1	HMIVV	hemagglutinin prec
16	83.5	9.8	639	2	A32935	protein P1 - Entam
17	83.5	9.8	1114	2	JH0284	125K surface antig
18	82.5	9.7	315	1	HNVZVT	hemagglutinin prec
19	82.5	9.7	1155	2	H71456	probable pyrolysin
20	81.5	9.6	550	1	HMIVS3	hemagglutinin prec
21	81	9.6	824	1	S50767	S-receptor kinase
22	80.5	9.5	473	1	A59200	acid phosphatase (
23	80	9.4	314	1	JQ1793	hemagglutinin prec
24	80	9.4	314	1	HNVZVW	hemagglutinin prec
25	80	9.4	385	2	S55524	neurokinin 3 recep
26	79.5	9.4	315	2	T37438	hemagglutinin - va
27	79	9.3	373	2	JH0155	pectin lyase (EC 4
28	79	9.3	1064	2	B86465	probable Protein k
29	78.5	9.3	389	2	AH3003	penicillin-binding
30	78.5	9.3	398	2	A99280	penicillin-binding
31	78.5	9.3	550	1	HMIV80	hemagglutinin prec
32	78.5	9.3	550	1	HMIV15	hemagglutinin prec
33	78	9.2	412	2	S72579	hypothetical prote
34	78	9.2	1305	2	H41662	150K mating aggreg
35	77.5	9.1	315	1	HNVZ4X	hemagglutinin prec
36	77.5	9.1	550	1	HMIVS2	hemagglutinin prec
37	77.5	9.1	550	1	HMIV77	hemagglutinin prec
38	77.5	9.1	550	1	HMIV89	hemagglutinin prec
39	77.5	9.1	550	2	JQ1153	hemagglutinin prec
40	77.5	9.1	566	1	HMIVHA	hemagglutinin prec
41	77	9.1	433	2	D90176	hypothetical prote
42	76.5	9.0	375	2	T00467	probable UDPglucos
43	76.5	9.0	550	2	JQ1154	hemagglutinin prec
44	76.5	9.0	550	2	JQ1155	hemagglutinin prec
45	76.5	9.0	566	1	HMIVH	hemagglutinin prec

ALIGNMENTS

RESULT 1
S56545
fimbrial protein fimH precursor, type 1 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 01-Mar-2002
C;Accession: S56545; B65246; S09563; A36967
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from
A;Reference number: S56314; MUID:95334362; PMID:7610040
A;Accession: S56545
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-300 <BUR>
A;Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97216.1; PID:g537161
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B65246
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-300 <BLAT>
A;Cross-references: GB:AE000502; GB:U00096; NID:g2367374; PIDN:AAC77276.1; PID:g17907
A;Experimental source: strain K-12, substrain MG1655
R;Klemm, P.; Christiansen, G.
Mol. Gen. Genet. 208, 439-445, 1987
A;Title: Three fim genes required for the regulation of length and mediation of adhes
A;Reference number: S07321; MUID:88038337; PMID:2890081
A;Accession: S09563
A;Molecule type: DNA
A;Residues: 1-196,'R',198-221,'H',223-300 <KLE>
A;Cross-references: EMBL:X05672; NID:g41463; PIDN:CAA29156.1; PID:g41466
R;Sokurenko, E.V.; Courtney, H.S.; Ohman, D.E.; Klemm, P.; Hasty, D.L.
J. Bacteriol. 176, 748-755, 1994
A;Title: FimH family of type 1 fimbrial adhesins: functional heterogeneity due to min
A;Reference number: A36967; MUID:94131954; PMID:7905476
A;Accession: A36967
A;Status: not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-78,'R',80-221,'H',223-300 <SOK>
A;Note: sequence extracted from NCBI backbone (NCBIP:143314)
C;Genetics:
A;Gene: fimH
C;Function:
A;Description: involved in longitudinal regulation and mannose-specific adhesion
A;Note: not necessary for the production of fimbriae
A;Note: controls length and number of fimbriae
C;Superfamily: fimbrial protein fimH
C;Keywords: fimbria

```
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-300/Product: fimbrial protein fimH, type 1 #status predicted <MAT>

  Query Match             98.3%;  Score 834;  DB 2;  Length 300;
  Best Local Similarity   98.1%;  Pred. No. 8.6e-70;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 2
G91288
hypothetical protein ECs5279 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: G91288
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G91288
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-300 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB38702.1; PID:g13364757; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs5279
C;Superfamily: fimbrial protein fimH

  Query Match             97.2%;  Score 824;  DB 2;  Length 300;
  Best Local Similarity   96.9%;  Pred. No. 7.2e-69;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTKNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 3
B86130
hypothetical protein fimH [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B86130
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B86130
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-300 <STO>
A;Cross-references: GB:AE005174; NID:g12519327; PIDN:AAG59502.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: fimH
C;Superfamily: fimbrial protein fimH

  Query Match             97.2%;  Score 824;  DB 2;  Length 300;
  Best Local Similarity   96.9%;  Pred. No. 7.2e-69;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: ||||||||||||||| ||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTKNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 4
A32801
fimbrial adhesin precursor, type 1 - Klebsiella pneumoniae
C;Species: Klebsiella pneumoniae
C;Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 18-Sep-1998
C;Accession: A32801
R;Gerlach, G.F.; Clegg, S.; Allen, B.L.
J. Bacteriol. 171, 1262-1270, 1989
A;Title: Identification and characterization of the genes encoding the type 3 and typ
A;Reference number: A32801; MUID:89155420; PMID:2563996
A;Accession: A32801
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-302 <GER>
A;Cross-references: GB:M24564
C;Superfamily: fimbrial protein fimH

  Query Match             89.2%;  Score 756;  DB 2;  Length 302;
  Best Local Similarity   86.3%;  Pred. No. 1.5e-62;
  Matches  139;  Conservative   12;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||:|||||||:|:|||||
Db      49 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGTSYPFP 108

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||:|| ||:|:|||||||  |||||||:|||: | |||||||||| |||||||| |||:
Db     109 TTTETARVIYDSRTDKPWPAVLYLTPVSTAGGVAITAGSLIAVLILHQTNNYNSDSFQFI 168

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||| ||: :||||:
Db     169 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSMAVPLTVH 209


RESULT 5
A64904
probable fimbrial protein b1502 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: A64904
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A64904
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-304 <BLAT>
A;Cross-references: GB:AE000247; GB:U00096; NID:g1787773; PIDN:AAC74575.1; PID:g17877
```

```
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: fimbrial protein fimH
C;Keywords: fimbria

  Query Match             45.3%;  Score 384;  DB 2;  Length 304;
  Best Local Similarity   47.9%;  Pred. No. 4.5e-28;
  Matches   78;  Conservative   30;  Mismatches   53;  Indels    2;  Gaps    2;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETI-TDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPF 59
           ||:  |||||||||  | | |||     ||:: | :|||: | | :: |:: ::  :|||
Db      48 PVIQPGQNLVVDLSQHISCWNDYGGWYDTDHINLVQGSAFAGSLQSYKGSLYWNNVTYPF 107

Qy      60 PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNS-DDFQ 118
           | |: |  :    :|  | |: ||:||| :|||:|||| :|| : : :     | :
Db     108 PLTTNTNVLDIGDKTPMPLPLKLYITPVGAAGGVVIKAGEVIARIHMYKIATLGSGNPRN 167

Qy     119 FVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           | ||| :||:||:|||| | :|:||| |||: ||  ||| ||
Db     168 FTWNIISNNNVVMPTGGCTVDSRNVTVDLPDFPGSAEIPLGVY 210


RESULT 6
C90892
probable adhesin [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: C90892
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90892
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-304 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35530.1; PID:g13361573; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2107
C;Superfamily: fimbrial protein fimH

  Query Match             45.3%;  Score 384;  DB 2;  Length 304;
  Best Local Similarity   47.9%;  Pred. No. 4.5e-28;
  Matches   78;  Conservative   29;  Mismatches   54;  Indels    2;  Gaps    2;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETI-TDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPF 59
           ||:  |||||||||  | | |||     ||:: | :|||: | | :: |:: ::  :|||
Db      48 PVIQPGQNLVVDLSQHISCWNDYGGWYDTDHINLVQGSAFAGSLQSYKGSLYWNNVTYPF 107

Qy      60 PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNS-DDFQ 118
           | |: |  :    :|  | |: ||:||| :|||:|||| :|| : : :     | :
Db     108 PLTTNTNVLDIGDKTPMPLPLKLYITPVGAAGGVVIKAGEVIARIHMYKIATLGSGNPRN 167

Qy     119 FVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           | ||| :|| ||:|||| | :|:||| |||: ||  ||| ||
Db     168 FTWNIISNNSVVMPTGGCTVDSRNVTVNLPDFPGSAEIPLGVY 210


RESULT 7
F85725
probable adhesin, FimH type protein Z2206 [imported] - Escherichia coli (strain O157:H7,
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: F85725
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85725
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-304 <STO>
A;Cross-references: GB:AE005174; NID:g12515169; PIDN:AAG56266.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z2206
C;Superfamily: fimbrial protein fimH

  Query Match             45.3%;  Score 384;  DB 2;  Length 304;
  Best Local Similarity   47.9%;  Pred. No. 4.5e-28;
  Matches   78;  Conservative   29;  Mismatches   54;  Indels    2;  Gaps    2;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETI-TDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPF 59
           ||:  |||||||||  | | |||     ||:: | :|||: | | :: |:: ::  :|||
Db      48 PVIQPGQNLVVDLSQHISCWNDYGGWYDTDHINLVQGSAFAGSLQSYKGSLYWNNVTYPF 107

Qy      60 PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNS-DDFQ 118
           | |: |  :    :|  | |: ||:||| :|||:|||| :|| : : :     | :
Db     108 PLTTNTNVLDIGDKTPMPLPLKLYITPVGAAGGVVIKAGEVIARIHMYKIATLGSGNPRN 167

Qy     119 FVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           | ||| :|| ||:|||| | :|:||| |||: ||  ||| ||
Db     168 FTWNIISNNSVVMPTGGCTVDSRNVTVNLPDFPGSAEIPLGVY 210


RESULT 8
C49233
S fimbrial adhesin minor subunit SfaH - Escherichia coli
C;Species: Escherichia coli
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C;Accession: C49233
R;Hacker, J.; Kestler, H.; Hoschutzky, H.; Jann, K.; Lottspeich, F.; Korhonen, T.K.
Infect. Immun. 61, 544-550, 1993
A;Title: Cloning and characterization of the S fimbrial adhesin II complex of an Esch
A;Reference number: A49233; MUID:93138776; PMID:8093693
A;Contents: O18:K1
A;Accession: C49233
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-299 <HAC>
A;Cross-references: GB:S53210; NID:g264033; PIDN:AAB25047.1; PID:g264036
A;Note: sequence extracted from NCBI backbone (NCBIN:123181, NCBIP:123186)
C;Superfamily: fimbrial protein fimH

  Query Match             24.4%;  Score 207;  DB 2;  Length 299;
  Best Local Similarity   32.6%;  Pred. No. 1.1e-11;
  Matches   56;  Conservative   22;  Mismatches   82;  Indels   12;  Gaps    5;

Qy       1 PVVNVGQNL-VVDLSTQIFCHN-DYPETITDYVTLQRGSAYGGVLS-------NFSGTVK 51
           |||   ::: |:|||  : | | |       ||: : :|| :   |        :|:
Db      34 PVVQYDKSISVLDLSQLVSCQNEDSTGQNYDYLKILKGSGFSPALDTKTYGRLDFTSRPT 93

Qy      52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQ--T 109
                |      :      |      ||:|  ||| |      | ||  | |:| | : :  |
Db      94 GYARQLPLQFDLQVTEAFYQYGVWKPFPAKLYLYPAPGVFGKVINNGDLLATLYVNKFST 153

Qy     110 NNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGS-VPIPLTV 160
             : :  | |  || ||| : || | ||: :| | || | |  | :||||
Db     154 KGQEAGERNFTWRFYATNDVHIQTGTCRVSSNNVKVDLPSYPGGPVTVPLTV 205


RESULT 9
S15927
sfaH protein precursor - Escherichia coli
C;Species: Escherichia coli
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 08-Oct-1999
C;Accession: S15927; S06195
R;Schmoll, T.; Hoschuetzky, H.; Morschhaeuser, J.; Lottspeich, F.; Jann, K.; Hacker,
Mol. Microbiol. 3, 1735-1744, 1989
A;Title: Analysis of genes coding for the sialic acid-binding adhesin and two other m
A;Reference number: S15925; MUID:90158121; PMID:2576095
A;Accession: S15927
A;Molecule type: DNA
A;Residues: 1-291 <SCH>
```

```
A;Cross-references: EMBL:X16664; NID:g42955; PIDN:CAA34654.1; PID:g42962
C;Genetics:
A;Gene: sfaH
C;Superfamily: fimbrial protein fimH
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-291/Product: sfaH protein #status predicted <MAT>

  Query Match             24.2%;  Score 205;  DB 2;  Length 291;
  Best Local Similarity   32.6%;  Pred. No. 1.7e-11;
  Matches   56;  Conservative   22;  Mismatches   82;  Indels   12;  Gaps    5;

Qy       1 PVVNVGQNL-VVDLSTQIFCHN-DYPETITDYVTLQRGSAYGGVLS-------NFSGTVK 51
           |||   ::: |:|||  : | | |       ||: : :|| :   |          :|:
Db      34 PVVEYDKSISVLDLSQLVSCQNEDSTGQNYDYLKILKGSGFSPALDTKTYGRLDFTSRPT 93

Qy      52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQ--T 109
                |     :     |     ||:|  ||| |     | ||  | |:| | : :  |
Db      94 GYARQLPLQFDLQVTEAFYQYGVWKPFPAKLYLYPEPGVFGKVINNGDLLATLYVNKFST 153

Qy     110 NNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGS-VPIPLTV 160
             : :  | |  || ||| : || | ||: :| | || | |  | :||||
Db     154 KGQEAGERNFTWRFYATNDVHIQTGTCRVSSNNVKVDLPSYPGGPVTVPLTV 205


RESULT 10
I76900
F1652 minor fimbrial chain H - Escherichia coli
C;Species: Escherichia coli
C;Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 08-Oct-1999
C;Accession: I76900; I80334
R;Harel, J.; Jacques, M.; Fairbrother, J.M.; Bosse, M.; Forget, C.
Microbiology 141, 221-228, 1995
A;Title: Cloning of determinants encoding F165(2) fimbriae from porcine septicaemic Esch
A;Reference number: I57111; MUID:95202083; PMID:7894716
A;Accession: I76900
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-299 <RES>
A;Cross-references: EMBL:U09804; NID:g967126; PIDN:AAA74946.1; PID:g967129
R;van Die, I.; Kramer, C.; Hacker, J.; Bergmans, H.; Jongen, W.; Hoekstra, W.
Res. Microbiol. 142, 653-658, 1991
A;Title: Nucleotide sequence of the genes coding for minor fimbrial subunits of the F1C
A;Reference number: I59446; MUID:92073661; PMID:1683712
A;Accession: I80334
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-299 <RE2>
A;Cross-references: GB:S68237; NID:g239708; PIDN:AAB20439.1; PID:g239711
C;Genetics:
A;Gene: focH
C;Superfamily: fimbrial protein fimH

  Query Match             23.5%;  Score 199.5;  DB 2;  Length 299;
  Best Local Similarity   33.3%;  Pred. No. 5.6e-11;
  Matches   54;  Conservative   24;  Mismatches   73;  Indels   11;  Gaps    6;

Qy      10 VVDLSTQIFCHN-DYPETITDYVTLQRGSAYGGVL-----SNFSGTVKYSGSSYPFPTTS 63
           |:||:  : | | |      ||: :::|: :   |         | : || |   |
Db      44 VLDLNQLVLCQNEDASGQNYDYLRVRQGTGFSPSLDAKTYGRLDFTNRLSGYSQTLPLQQ 103

Qy      64 ET-PRVVY-NSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQ--TNNYNSDDFQF 119
           :| |   |      ||:|  :|| |     | :| || |:| : : :  |    : :  |
Db     104 DTKPTEAYWQYGVWKPFPAKMYLYPEPGVFGKLIHAGELVATVYVNKFSTMGQEAGERNF 163

Qy     120 VWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGS-VPIPLTV 160
            |  || ||| : || | ||: :| | || | |  | :||||
Db     164 TWRFYATNDVYIQTGTCRVSSNNVKVDLPSYPGGPVTVPLTV 205


RESULT 11
HMIV98
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: F27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: F27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16742; NID:g324091
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             10.2%;  Score 86.5;  DB 1;  Length 550;
  Best Local Similarity   25.0%;  Pred. No. 3.6;
  Matches   37;  Conservative   18;  Mismatches   58;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:         |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPNSGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |               : |:||   || :   || | |   ::| ::         |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
           |  :||  || |    |:    || |
Db     194 LYVQASGGVTVSTRRSQQTIIPNIGSRP 221


RESULT 12
HNVZVV
hemagglutinin precursor - vaccinia virus
C;Species: vaccinia virus
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
C;Accession: A27789
R;Shida, H.
Virology 150, 451-462, 1986
A;Title: Nucleotide sequence of the vaccinia virus hemagglutinin gene.
A;Reference number: A27789; MUID:86181588; PMID:3008418
A;Accession: A27789
A;Molecule type: DNA
A;Residues: 1-315 <SHI>
A;Cross-references: GB:M14783; GB:M14130; NID:g335633; PIDN:AAA48251.1; PID:g335634
C;Superfamily: vaccinia virus hemagglutinin; immunoglobulin homology
C;Keywords: glycoprotein; hemagglutinin; late protein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-315/Product: hemagglutinin #status predicted <HNN>
F;27-105/Domain: immunoglobulin homology <IMM>
F;280-302/Domain: transmembrane #status predicted <TMM>
F;303-315/Domain: intracellular #status predicted <INT>
F;37,69,112,161,254/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             10.1%;  Score 85.5;  DB 1;  Length 315;
  Best Local Similarity   25.5%;  Pred. No. 2.2;
  Matches   36;  Conservative   17;  Mismatches   37;  Indels   51;  Gaps    7;

Qy      24 PETITDYVTLQRGSAYGGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNS 72
           || ||| |             : : || |          || |    || ||| : :
Db     171 PEPITDNV-----------EDHTDTVTYTSDSINTVSASSGES----TTDETPEPITDK 214

Qy      73 RTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVP 132
           |      :  | ||:: |:|             |    :||   ::: | :|| | |
Db     215 EEDHTVTDTVSYTTVSTSSGIV----------TTKSTTDDAD-LYDTYNDNDTVPP 259
```

```
Qy    133 T--GGCDVSARDVTVTLPDYR 151
          |  ||        | :: :|:
Db    260 TTVGGS-------TTSISNYK 273


RESULT 13
HMIV33
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: C27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: C27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             10.1%;  Score 85.5;  DB 1;  Length 550;
  Best Local Similarity   25.0%;  Pred. No. 4.4;
  Matches   37;  Conservative   18;  Mismatches   58;  Indels   35;  Gaps    6;

Qy     21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
          | ||  : || :|:   |  | |        ::|  :  ||:         |   : ||
Db     96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPASGFFSRL 151

Qy     75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
          :  |              : |:||   || :   || | |   ::| ::         |
Db    152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy    128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
            |  :||  || |    |:    || |
Db    194 LYVQASGGVTVSTRRSQQTIIPNIGSRP 221


RESULT 14
F69195
cell surface glycoprotein (s-layer protein) - Methanobacterium thermoautotrophicum (stra
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Dec-1997
C;Accession: F69195
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: F69195
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1797 <MTH>
A;Cross-references: GB:AE000666
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH716

  Query Match              9.9%;  Score 84;  DB 2;  Length 1797;
  Best Local Similarity   23.9%;  Pred. No. 27;
  Matches   44;  Conservative   25;  Mismatches   61;  Indels   54;  Gaps    9;

Qy     26 TITDYVTLQRGSAY-------------------------GGVLSNFSGTVKYSGSSYP 58
          |  :|:|: :|  :                         | |  |::  :| :| :
Db   1465 TKENYITVYQGPQFTLENLTVIPTSGVAPLNITVSANLTNSGDVAGNYTAQLKVNGQTVA 1524

Qy     59 FPTTS---------ETPRV----VYNSRTDKPWPVAL-YLTPVS-SAGGLVIKAGSLIAV 103
            | :         |  ||    |||    |   |||:  | | : : | | :   | :|
Db   1525 TKTIAVPAGQTVSVEFSRVLTAGVYNVTIDDLAPVAVTVLKPANITVGNLTVTPKSGVAP 1584

Qy    104 LILRQTNNY-NSDDF------QFVWN--IYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
          | :  | |  |: |        ||  |     |   : || ||  | : : |:|     |:
Db   1585 LNITVTANLRNTGDLAGNYTAQFKVNGQTVATKTIAVPAGGVGVVSFNYTLT---SAGNY 1641

Qy    155 PIPL 158
          |: :
Db   1642 PVAI 1645


RESULT 15
HMIVV
hemagglutinin precursor - influenza A virus (strain A/Victoria/3/75)
C;Species: influenza A virus
C;Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 28-May-1999
C;Accession: A90794; A04050; A92790
R;Min Jou, W.; Verhoeyen, M.; Devos, R.; Saman, E.; Fang, R.; Huylebroeck, D.; Fiers,
Cell 19, 683-696, 1980
A;Title: Complete structure of the hemagglutinin gene from the human influenza A/Vict
A;Reference number: A90794; MUID:80155186; PMID:6153930
A;Accession: A90794
A;Molecule type: genomic RNA
A;Residues: 1-567 <MIN>
A;Cross-references: GB:V01098; GB:J02172; GB:M55060; NID:g60784; PIDN:CAA24281.1; PID
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-345/Product: hemagglutinin HA1 #status predicted <HA1>
F;347-567/Product: hemagglutinin HA2 #status predicted <HA2>
F;537-553/Domain: transmembrane #status predicted <TM1>
F;31-483,69-294,81-93,156-490,298-322/Disulfide bonds: #status predicted
F;556,563,566/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match              9.8%;  Score 83.5;  DB 1;  Length 567;
  Best Local Similarity   26.4%;  Pred. No. 7;
  Matches   39;  Conservative   17;  Mismatches   57;  Indels   35;  Gaps    7;

Qy     21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
          | ||  : || :|:   |  | |       |::|  :  |||          |   : ||
Db    113 NCYPYDVPDYASLRSLVASSGTLEFINEGFNWTGVTQNGGSS----ACKRGPDSGFFSRL 168

Qy     75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
          :  |   ||            |:||   |  :   || |||   ::| ::         |
Db    169 N--W---LY------------KSGSTYPVQNVTMPNNDNSDKL-YIWGVHHPSTDKEQTN 210

Qy    128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
            |  :|   || :    |:    || |
Db    211 LYVQASGKVTVSTKRSQQTIIPNVGSRP 238


RESULT 16
A32935
protein P1 - Entamoeba histolytica (fragment)
C;Species: Entamoeba histolytica
C;Date: 20-Dec-1989 #sequence_revision 20-Dec-1989 #text_change 24-Feb-1995
C;Accession: A32935
R;Tannich, E.; Horstmann, R.D.; Knobloch, J.; Arnold, H.H.
Proc. Natl. Acad. Sci. U.S.A. 86, 5118-5122, 1989
A;Title: Genomic DNA differences between pathogenic and nonpathogenic Entamoeba histo
A;Reference number: A32935; MUID:89296955; PMID:2544890
A;Accession: A32935
A;Status: preliminary
A;Molecule type: mRNA
```

```
A;Residues: 1-639 <TAN>

  Query Match          9.8%;  Score 83.5;  DB 2;  Length 639;
  Best Local Similarity   23.6%;  Pred. No. 8.1;
  Matches   38;  Conservative   32;  Mismatches   58;  Indels   33;  Gaps    7;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSY-PF 59
           |: ::| :| :: :|| :  ||         | : || :||::   | |  :: | |  |
Db      28 PIDSIGFDLGLN-TTQPYIIND---------TFKIGSPFGGMIYLRSDTT-FTNSFYVTF 76

Qy      60 PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQF 119
                | : ||  |:: |   |      :| | | :  :    |:|   |  : :| |:
Db      77 SNVGRAPIINYNITTNEEWNSVL-----RNAPGNVAEIRTPGNRLVLTSRNIRSLEDAQY 131

Qy     120 VWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTV 160
           : : :           :|  :  ||| :      ||:|:
Db     132 ISDFWLK----------AISISNYAVTLEN------IPITL 156


RESULT 17
JH0284
125K surface antigen M17 precursor - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: JH0284
R;Edman, U.; Meraz, M.A.; Rausser, S.; Agabian, N.; Meza, I.
J. Exp. Med. 172, 879-888, 1990
A;Title: Characterization of an immuno-dominant variable surface antigen from pathogenic
A;Reference number: JH0284; MUID:90354789; PMID:1696956
A;Accession: JH0284
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1114 <EDM>
A;Experimental source: strain HM1 IMSS

  Query Match          9.8%;  Score 83.5;  DB 2;  Length 1114;
  Best Local Similarity   23.6%;  Pred. No. 16;
  Matches   38;  Conservative   32;  Mismatches   58;  Indels   33;  Gaps    7;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSY-PF 59
           |: ::| :| :: :|| :  ||         | : || :||::   | |  :: | |  |
Db     214 PIDSIGFDLGLN-TTQPYIIND---------TFKIGSPFGGMIYLRSDTT-FTNSFYVTF 262

Qy      60 PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQF 119
                | : ||  |:: |   |      :| | | :  :    |:|   |  : :| |:
Db     263 SNVGRAPIINYNITTNEEWNSVL-----RNAPGNVAEIRTPGNRLVLTSRNIRSLEDAQY 317

Qy     120 VWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTV 160
           : : :           :|  :  ||| :      ||:|:
Db     318 ISDFWLK----------AISISNYAVTLEN------IPITL 342


RESULT 18
HNVZVT
hemagglutinin precursor - vaccinia virus (strain Tiantan)
C;Species: vaccinia virus
C;Date: 30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change 16-Jul-1999
C;Accession: JL0108
R;Jin, D.; Li, Z.; Jin, Q.; Yuwen, H.; Hou, Y.
J. Exp. Med. 170, 571-576, 1989
A;Title: Vaccinia virus hemagglutinin. A novel member of the immunoglobulin superfamily.
A;Reference number: JL0108; MUID:89328331; PMID:2754392
A;Accession: JL0108
A;Molecule type: mRNA
A;Residues: 1-315 <JIN>
A;Cross-references: GB:X15709; GB:M57773; NID:g61313; PIDN:CAA33740.1; PID:g61314
C;Superfamily: vaccinia virus hemagglutinin; immunoglobulin homology
C;Keywords: glycoprotein; hemagglutinin; late protein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-315/Product: hemagglutinin #status predicted <MAT>
F;27-105/Domain: immunoglobulin homology <IMM>
F;280-302/Domain: transmembrane #status predicted <TMM>
F;303-315/Domain: intracellular #status predicted <INT>
F;37,69,112,161,254/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match          9.7%;  Score 82.5;  DB 1;  Length 315;
  Best Local Similarity   23.8%;  Pred. No. 4.2;
  Matches   38;  Conservative   22;  Mismatches   43;  Indels   57;  Gaps    8;

Qy       4 NVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKY---------- 52
           |:  : | :::|        || ||| |             : : || |
Db     159 NLNCSSVFEIAT--------PEPITDNV-----------EDHTDTVTYTSDSINTVSAT 198

Qy      53 SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNY 112
           || |    || |||  : :   |      :  | ||:: |:|                 |
Db     199 SGES----TTDETPEPITDKEEDHTVTDTVSYTTVSTSSGIV-------------TTKS 240

Qy     113 NSDDFQFVWNIYANNDVVVPTG-GCDVSARDVTVTLPDYR 151
            :||   ::: | :|| |  |  ||       | :: :|:
Db     241 TTDDAD-LYDTYNDNDTVPSTTVGCS------TTSISNYK 273


RESULT 19
H71456
probable pyrolysin (EC 3.4.-.-) homolog PH0310 [similarity] - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: H71456
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophili
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: H71456
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1155 <KAW>
A;Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29383.1; PID:g3256700
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBa
C;Genetics:
A;Gene: PH0310
C;Keywords: hydrolase

  Query Match          9.7%;  Score 82.5;  DB 2;  Length 1155;
  Best Local Similarity   24.5%;  Pred. No. 21;
  Matches   46;  Conservative   24;  Mismatches   59;  Indels   59;  Gaps   11;

Qy      14 STQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPR------ 67
           :| :: || | :   ||::      |  :||| |    |   ::|||   |:|
Db     566 NTIVWFHN-YLDFWIDYIS---DKEYNAILSNLSLLESYLPQAHPFP---ESPMAFDVSQ 618

Qy      68 ----VVY--NSRTDKPWPVALYLTP----------------VSSAGGLVIKAGSLIAVL 104
              :||    :   |  |  |: |                   : | ||     |||
Db     619 LSDYLVYYMEDKNSTPTEVTFYVQPKDILPDNFTGELYNFTLLISIGGEYKVLGSL---- 674

Qy     105 ILRQTNNYNSDDFQFV----------WNIYANNDVVVPTGGCDVSARDVTVTLPDYR--- 151
                | :  |  :|          ::: | |:| |  |  :||   : : | | |
Db     675 ----EGNVSIPDGGYVTVPLEIPVVSYDV-APNNVTVQYGNFNVSIYSI-IPLEDARIVI 728

Qy     152 GSVPIPLT 159
           |:   |||
Db     729 GNESYPLT 736


RESULT 20
HMIVS3
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C;Species: influenza A virus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C;Accession: B29971
R;Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A;Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in Chin
A;Reference number: A94370; MUID:88101364; PMID:3336940
A;Accession: B29971
```

```
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M19057; NID:g324210
A;Note: the sequence in GenBank entry FLAHAPB, release 106, (PID:g324211) differs from t
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             9.6%;  Score 81.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 10;
  Matches   36;  Conservative   19;  Mismatches   58;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:         |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPANGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
             :  |               : |:||   || :   || |||   ::| ::         |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNSDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || :    |:    || |
Db     194 LYVQASGRVTVSTKRSQQTMIPNAGSRP 221


RESULT 21
S50767
S-receptor kinase (EC 2.7.1.-) homolog precursor - rice
C;Species: Oryza sativa (rice)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S50767
R;Zhao, Y.; Feng, X.H.; Watson, J.C.; Bottino, P.J.; Kung, S.D.
Plant Mol. Biol. 26, 791-803, 1994
A;Title: Molecular cloning and biochemical characterization of a receptor-like serine/th
A;Reference number: S50767; MUID:95093022; PMID:7999995
A;Accession: S50767
A;Molecule type: mRNA
A;Residues: 1-824 <ZHA>
A;Cross-references: EMBL:L27821; NID:g450299; PIDN:AAA33915.1; PID:g450300
C;Superfamily: S-receptor kinase; protein kinase homology; S-locus-specific glycoprotein
C;Keywords: ATP; glycoprotein; magnesium; phosphotransferase; receptor; serine/threonine
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-824/Product: S-receptor kinase homolog PK10 #status predicted <MAT>
F;435-458/Domain: transmembrane #status predicted <TMM>
F;500-777/Domain: protein kinase homology <KIN>
F;506-514/Region: protein kinase ATP-binding motif
F;60,72,189,235,262,366,377,407/Binding site: carbohydrate (Asn) (covalent) #status pred
F;528,543,625,627/Active site: Lys, Glu, Asp, Lys #status predicted
F;630,634/Binding site: magnesium (Asn, Asp) #status predicted

  Query Match             9.6%;  Score 81;  DB 1;  Length 824;
  Best Local Similarity   23.4%;  Pred. No. 19;
  Matches   39;  Conservative   29;  Mismatches   57;  Indels   42;  Gaps    9;

Qy       7 QNLVV--DLSTQIFCHNDYPETITDYVTLQRGSAY--GGVLSNFSGTVK---------- 51
           |::|:  | |: ::    :|   ||  ||  |  :  |  | : | ||:
Db     145 QSVVLGKDASSPLWQSFSHP---TD--TLLSGQNFIEGMTLMSKSNTVQNMTYTLQIKSG 199

Qy      52 ----YSGSSYPFP--TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLI 105
               |:|   | |  :  :  |:: |   |: :|  |     |::     ::|||:: |:
Db     200 NMILYAGFETPQPYWSAQQDSRIIVNKNGDRIYPANL----SSASWSFYDQSGSLLSQLV 255

Qy     106 LRQTNN-------YNSDDFQFVWNIYANN-----DVVVPTGGCDVSA 140
           : | |          ||     : :   |      : ||   ||: |
Db     256 IAQENANATLSAVLGSDGLIAFYMLQGGNGKSKFSITVPADSCDMPA 302


RESULT 22
A59200
acid phosphatase (EC 3.1.3.2) purple 1, precursor [validated] - sweet potato
N;Alternate names: purple acid phosphatase (PAP)
C;Species: Ipomoea batatas (sweet potato)
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 24-Aug-2001
C;Accession: A59200; C59200
R;Schenk, G.; Ge, Y.; Carrington, L.E.; Wynne, C.J.; Searle, I.R.; Carroll, B.J.; Ham
Arch. Biochem. Biophys. 370, 183-189, 1999
A;Title: Binuclear metal centers in plant purple acid phosphatases: Fe-Mn in sweet po
A;Reference number: Z25293; MUID:99441212; PMID:10510276
A;Accession: A59200
A;Molecule type: mRNA
A;Residues: 1-473 <KUR>
A;Cross-references: GB:AF200825; NID:g6635440; PIDN:AAF19821.1; PID:g6635441
A;Accession: C59200
A;Molecule type: protein
A;Residues: 39-61 <KU2>
C;Comment: Metal content is iron and manganese, with small amounts of iron and copper
C;Complex: homodimer
C;Function:
A;Description: catalyzes the hydrolysis of phosphate monoesters [validated, MUID:9944
A;Note: active on a variety of activated and unactivated phosphate ester and anhydrid
C;Superfamily: kidney bean purple acid phosphatase; phosphoesterase core homology
C;Keywords: glycoprotein; homodimer; iron; manganese; metalloprotein; phosphoric mono
F;39-473/Product: purple acid phosphatase #status experimental <MAT>
F;166-240/Domain: phosphoesterase core homology <PEC>
F;118,180,311,433/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;172,201,204,362/Binding site: iron (Asp, Asp, Tyr, His) #status predicted
F;201,238,323,360/Binding site: manganese (Asp, Asn, His, His) #status predicted
F;239,333/Active site: His #status predicted
F;382/Disulfide bonds: interchain #status predicted

  Query Match             9.5%;  Score 80.5;  DB 1;  Length 473;
  Best Local Similarity   37.9%;  Pred. No. 11;
  Matches   22;  Conservative   12;  Mismatches   19;  Indels    5;  Gaps    2;

Qy      30 YVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPV 87
             :  ::| ||:  |||::|| ||||     |  |||  :|   :|:: || : |   |:
Db     274 WYAIKRASAHIIVLSSYSGFVKYSPQYKWF--TSELEKV---NRSETPWLIVLVHAPL 326


RESULT 23
JQ1793
hemagglutinin precursor - vaccinia virus (strain WR)
N;Alternate names: SalG1R protein
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jun-2000
C;Accession: JQ1793
R;Smith, G.L.; Chan, Y.S.; Howard, S.T.
J. Gen. Virol. 72, 1349-1376, 1991
A;Title: Nucleotide sequence of 42kbp of vaccinia virus strain WR from near the right
A;Reference number: JQ1767; MUID:91259063; PMID:2045793
A;Accession: JQ1793
A;Molecule type: DNA
A;Residues: 1-314 <SMI>
A;Cross-references: DDBJ:D11079; NID:g222717; PIDN:BAA01829.1; PID:g222744
C;Superfamily: vaccinia virus hemagglutinin; immunoglobulin homology
C;Keywords: glycoprotein; hemagglutinin; late protein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-314/Product: hemagglutinin #status predicted <HEG>
F;27-105/Domain: immunoglobulin homology <IMM>
F;279-301/Domain: transmembrane #status predicted <TMM>
F;37,69,112,253/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             9.4%;  Score 80;  DB 1;  Length 314;
  Best Local Similarity   22.7%;  Pred. No. 7.1;
  Matches   42;  Conservative   21;  Mismatches   48;  Indels   74;  Gaps   10;

Qy       2 VVNVGQNLVVDLSTQIFCHN--------DY-------------PETITDYVTLQRGSAY 39
             :||      :|:      |:        ||              || ||| |
```

```
Db    127 IVNTDSESTIDIILSGSTHSPETSSKKPDYIDNSNCSSVFEIATPEPITDNV-------- 178

Qy     40 GGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
          : : || |           || |    || ||| :   : |      :  | ||
Db    179 ----EDHTDTVTYTSDSINTVSASSGES----TTDETPEPI-TDKEDHTVTDTVSYTTVS 229

Qy     89 SAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPT--GGCDVSARDVTVT 146
          :: |:|             |    :||   ::: | :|| | ||  ||        | :
Db    230 TSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPPTTVGGS-------TTS 267

Qy    147 LPDYR 151
          : :|:
Db    268 ISNYK 272


RESULT 24
HNVZVW
hemagglutinin precursor - raccoonpox virus (strain WR)
C;Species: raccoonpox virus
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
C;Accession: B43381
R;Cavallaro, K.F.; Esposito, J.J.
Virology 190, 434-439, 1992
A;Title: Sequences of the raccoon poxvirus hemagglutinin protein.
A;Reference number: A43381; MUID:92410621; PMID:1529542
A;Accession: B43381
A;Molecule type: DNA
A;Residues: 1-314 <CAV>
A;Cross-references: GB:M93956; NID:g335637; PIDN:AAA48252.1; PID:g335638
C;Superfamily: vaccinia virus hemagglutinin; immunoglobulin homology
C;Keywords: glycoprotein; hemagglutinin; late protein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-314/Product: hemagglutinin #status predicted <HEG>
F;27-105/Domain: immunoglobulin homology <IMM>
F;280-296/Domain: transmembrane #status predicted <TMM>
F;34-103/Disulfide bonds: #status predicted
F;37,69,112,161,253/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             9.4%;  Score 80;  DB 1;  Length 314;
  Best Local Similarity   22.7%;  Pred. No. 7.1;
  Matches   42;  Conservative   21;  Mismatches   48;  Indels   74;  Gaps   10;

Qy      2 VVNVGQNLVVDLSTQIFCHN--------DY--------------PETITDYVTLQRGSAY 39
          :||      :|:      |:         ||              || ||| |
Db    127 IVNTDSESTIDIILSGSTHSPETSSKKPDYIDNSNCSSVFEIATPEPITDNV-------- 178

Qy     40 GGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
          : : || |           || |    || ||| :   : |      :  | ||
Db    179 ----EDHTDTVTYTSDSINTVSASSGES----TTDETPEPI-TDKEDHTVTDTVSYTTVS 229

Qy     89 SAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPT--GGCDVSARDVTVT 146
          :: |:|             |    :||   ::: | :|| | ||  ||        | :
Db    230 TSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPPTTVGGS-------TTS 267

Qy    147 LPDYR 151
          : :|:
Db    268 ISNYK 272


RESULT 25
S55524
neurokinin 3 receptor - mouse (fragment)
N;Alternate names: neuromedin K receptor; NK-3 receptor
C;Species: Mus musculus (house mouse)
C;Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 19-May-2000
C;Accession: S55524; I73045
R;Maroteaux, L.
submitted to the EMBL Data Library, June 1995
A;Reference number: S55524
A;Accession: S55524
A;Molecule type: mRNA
A;Residues: 1-385 <MAR>
A;Cross-references: EMBL:X87823; NID:g861055; PIDN:CAA61088.1; PID:g861056
R;Cook, G.A.; Elliott, D.; Metwali, A.; Blum, A.M.; Sandor, M.; Lynch, R.; Weinstock,
J. Immunol. 152, 1830-1835, 1994
A;Title: Molecular evidence that granuloma T lymphocytes in murine schistosomiasis ma
A;Reference number: I56216; MUID:94165478; PMID:8120392
A;Accession: I73045
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 103-197,'S',199-266,'P',268-328 <COO>
A;Cross-references: GB:L27827; NID:g450288; PIDN:AAA17893.1; PID:g480780
C;Superfamily: neurokinin 1 receptor

  Query Match             9.4%;  Score 80;  DB 2;  Length 385;
  Best Local Similarity   30.3%;  Pred. No. 9.1;
  Matches   30;  Conservative   15;  Mismatches   22;  Indels   32;  Gaps    6;

Qy     27 ITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK----PWPVAL 82
          ||::: || |        |||       |:   | ||: |  | :: |::     | :||
Db     30 ITEWLALQAG--------NFS-------SALGLPVTSQAPSQVRDNLTNQFVQPSWRIAL 74

Qy     83 YLTPVSSAGGLVIKA---GSLIAVLI------LRQTNNY 112
          :    | | |||:     |:|| : |      :|   ||
Db     75 W----SLAYGLVVAVAVFGNLIVIWIILAHKRMRTVTNY 109


RESULT 26
T37438
hemagglutinin - vaccinia virus (strain Ankara)
C;Species: vaccinia virus
A;Variety: strain Ankara
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
C;Accession: T37438
R;Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F.
submitted to the EMBL Data Library, March 1997
A;Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) st
A;Reference number: Z20877
A;Accession: T37438
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-315 <ANT>
A;Cross-references: EMBL:U94848; PIDN:AAB96543.1
A;Experimental source: strain Ankara
C;Genetics:
A;Note: MVA165R
C;Superfamily: vaccinia virus hemagglutinin; immunoglobulin homology

  Query Match             9.4%;  Score 79.5;  DB 2;  Length 315;
  Best Local Similarity   22.2%;  Pred. No. 7.9;
  Matches   41;  Conservative   21;  Mismatches   50;  Indels   73;  Gaps    9;

Qy      2 VVNVGQNLVVDLSTQIFCHN--------DY--------------PETITDYVTLQRGSAY 39
          :||      :|:      |:         ||              || ||| |
Db    127 IVNTDSESTIDIILSGSTHSPETSSEKPDYIDNSNCSSVFEIATPEPITDNV-------- 178

Qy     40 GGVLSNFSGTVKY----------SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
          : : || |           || |    || ||| : :   |      :  | ||
Db    179 ----EDHTDTVTYTSDSINTVSASSGES----TTDETPEPITDKEEDHTVTDTVSYTTVS 230

Qy     89 SAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPT--GGCDVSARDVTVT 146
          :: |:|             |    :||   ::: | :|| |  |  ||        | :
Db    231 TSSGIV-------------TTKSTTDDAD-LYDTYNDNDTVPSTTVGGS-------TTS 268

Qy    147 LPDYR 151
          : :|:
Db    269 ISNYK 273


RESULT 27
JH0155
pectin lyase (EC 4.2.2.10) D precursor - Aspergillus niger
C;Species: Aspergillus niger
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 21-Jul-2000
C;Accession: JH0155; PS0295
R;Gysler, C.; Harmsen, J.A.M.; Kester, H.C.M.; Visser, J.; Heim, J.
```

```
Gene 89, 101-108, 1990
A;Title: Isolation and structure of the pectin lyase D-encoding gene from Aspergillus ni
A;Reference number: JH0155; MUID:90323592; PMID:2373363
A;Accession: JH0155
A;Molecule type: DNA
A;Residues: 1-373 <GYS1>
A;Cross-references: GB:M55657; GB:M33035; NID:g166515; PIDN:AAA32701.1; PID:g166516
A;Experimental source: strain N756
A;Accession: PS0295
A;Molecule type: protein
A;Residues: 20-30 <GYS2>
C;Genetics:
A;Gene: pelD
A;Introns: 68/2; 116/2; 158/2; 190/3
C;Keywords: carbon-oxygen lyase; glycoprotein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-373/Product: pectin lyase D #status experimental <MAT>
F;128,274,348/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             9.3%;  Score 79;  DB 2;  Length 373;
  Best Local Similarity   23.6%;  Pred. No. 11;
  Matches   38;  Conservative   25;  Mismatches   72;  Indels   26;  Gaps    7;

Qy       9 LVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGV-LSNFSGTVKYSGSSYPFPTTSETPR 67
           |  |  :::   |:|    :||     |  |  | |   |  | :|| :| : |:   |:
Db     200 LGTDADSRVSITNNYINGESDYSATCDGHHYWNVYLDGSSDKVTFSG-NYLYKTSGRAPK 258

Qy      68 V-------VYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIA-VLILRQTNN-----YNS 114
           |       :||:  :     |  :     :|| |:  |:  : |  : :|:      ::|
Db     259 VQDNTYLHIYNNYWENNSGHAFEI----GSGGYVLAEGNYFSNVDTVLETDTFEGALFSS 314

Qy     115 DDFQFVWNIYANNDVVVPTGGCDVSARDVTV-------TLP 148
           |       |     |  | |::    ||        |||
Db     315 DSASSTCESYIGRSCVANVNGGDLTGTSTTVLSNLSGDTLP 355


RESULT 28
B86465
probable Protein kinase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: B86465
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: B86465
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1064 <STO>
A;Cross-references: GB:AE005172; NID:g10086466; PIDN:AAG12526.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

  Query Match             9.3%;  Score 79;  DB 2;  Length 1064;
  Best Local Similarity   21.2%;  Pred. No. 40;
  Matches   45;  Conservative   21;  Mismatches   58;  Indels   88;  Gaps   10;

Qy      11 VDLSTQIFCHNDYPETITDYVTLQ-----RGSAYGGV---------------LSNFSGTV 50
           :|||   |  |  ||| :|   ||       | :| :               :|||| :
Db     602 LDLSYNTFTGN-IPETFSDLTQLQSLDLSSNSLHGDIKVLGSLTSLASLNISCNNFSGPI 660

Qy      51 KYSGSSYPFPTTSETPRVVYNSR------------------TDKPWPVALYLTPVSSAG 91
             | ||  |  |    : |:                     |  |||
Db     661 ----PSTPFFKTISTTSYLQNTNLCHSLDGITCSSHTGQNNGVKSPKIVAL--------- 707

Qy      92 GLVIKAGSLIAV-----LILRQTNNY----------NSDDFQFVWNIYANNDVVVPTGG 135
           || |   ||:     ||||  : |           ::|| : |         :|
Db     708 TAVILASITIAILAAWLLILRNNHLYKTSQNSSSSPSTAEDFSYPWTF-------IPFQK 760

Qy     136 CDVSARDVTVTLPD-----------YRGSVP 155
           ::  ::  :| |            |:  :|
Db     761 LGITVNNIVTSLTDENVIGKGCSGIVYKAEIP 792


RESULT 29
AH3003
penicillin-binding protein dacf precursor dacF [imported] - Agrobacterium tumefaciens
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AH3003
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AH3003
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-389 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44446.1; PID:g17742048; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: dacF
A;Map position: linear chromosome

  Query Match             9.3%;  Score 78.5;  DB 2;  Length 389;
  Best Local Similarity   24.2%;  Pred. No. 13;
  Matches   43;  Conservative   26;  Mismatches   64;  Indels   45;  Gaps    8;

Qy       1 PVVNVGQNLVVDLST-QIFCHND-----YPETITDYVTLQRGSAYGGVLSNFSGTVKYSG 54
           || |    :|||: | ::  | :     || ::|  :|     ||       :| :
Db      47 PVANANPKMVVDVKTGKVISHQEAFRKWYPASLTKLMT-----AYIAFSQMKAGKI---- 97

Qy      55 SSYPFPTTSETPRVVYNSRTDKPWPVALY-----LTPVSSAGGLVIKAGSLIAVLILRQT 109
             : :|  |:     |:|     :     ||  |:   |:||: : ||| |  :|
Db      98 -------SPQTEVVMSKKAADQPASKMYFKPGQKLTMDSALKLLLIKSANDIAVAI-AET 149

Qy     110 NNYNSDDFQFVWNIYAN----------NDVVVPTGGCDVSARDVTV-------TLPDY 150
             ||:|   |  |.           |   :|  |   :|||: :         |:|
Db     150 IGGTSDNFVAQMNAQAQRLGMSSTHYVNPNGLPGKGQYTTARDLALLALIIKREFPEY 207


RESULT 30
A99280
penicillin-binding protein dacF precursor rp389 [imported] - Agrobacterium tumefacien
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: A99280
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A;Reference number: A97359; PMID:11743194
A;Accession: A99280
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-398 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89763.1; PID:g15159685; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_2387
A;Map position: linear chromosome

  Query Match             9.3%;  Score 78.5;  DB 2;  Length 398;
  Best Local Similarity   24.2%;  Pred. No. 13;
  Matches   43;  Conservative   26;  Mismatches   64;  Indels   45;  Gaps    8;
```

```
Qy       1 PVVNVGQNLVVDLST-QIFCHND-----YPETITDYVTLQRGSAYGGVLSNFSGTVKYSG 54
           || |    :|||: | ::  |:     || ::|  :|     ||       :| :
Db      56 PVANANPKMVVDVKTGKVISHQEAFRKWYPASLTKLMT-----AYIAFSQMKAGKI---- 106

Qy      55 SSYPFPTTSETPRVVYNSRTDKPWPVALY-----LTPVSSAGGLVIKAGSLIAVLILRQT 109
                : :|  |:     |:|     :     ||  |:   |:||: : |||  |  :|
Db     107 -------SPQTEVVMSKKAADQPASKMYFKPGQKLTMDSALKLLLIKSANDIAVAI-AET 158

Qy     110 NNYNSDDFQFVWNIYAN----------NDVVVPTGGCDVSARDVTV-------TLPDY 150
             ||:|   |  |           |   :|  |   :|||: :         |:|
Db     159 IGGTSDNFVAQMNAQARLGMSSTHYVNPNGLPGKGQYTTARDLALLALIIKREFPEY 216


RESULT 31
HMIV80
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: B27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: B27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16738; NID:g324083
A;Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             9.3%;  Score 78.5;  DB 1;  Length 550;
  Best Local Similarity   23.9%;  Pred. No. 20;
  Matches   34;  Conservative   23;  Mismatches   62;  Indels   23;  Gaps    5;

Qy      21 NDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPV 80
           | ||  : || :|:        |   |||:|:    : :   ::       |   |  |
Db      96 NCYPYDVPDYASLRS-------LVASSGTLKFITEGFTWTGVTQNG----GSNACKRGPA 144

Qy      81 ALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANNDVVVPT 133
           : : : ::     : |:||   || :   || | |   ::| ::         |  |  :
Db     145 SGFFSRLN----WLTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTNLYVQAS 199

Qy     134 GGCDVSARDVTVTLPDYRGSVP 155
           |   || |    |:    || |
Db     200 GRVTVSTRRSQQTIIPNIGSRP 221


RESULT 32
HMIV15
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: G27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: G27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16743; NID:g324093; PIDN:AAA43149.1; PID:g324094
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             9.3%;  Score 78.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 20;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |         ::|  :  ||:         |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPNSGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |               : |:||   || :   || | |   ::| ::         |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || |    |:    || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 33
S72579
hypothetical protein C35D10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 16-Jul-1999
C;Accession: S72579
R;Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto
ulton, L.; Gardner, A.; Green, P.; Hawkins, T.; Hillier, L.; Jier, M.; Johnston, L.;
, B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Rifken, L.; Roopra, A.; Saunders, D.
submitted to the EMBL Data Library, February 1995
A;Authors: Shownkeen, R.; Smaldon, N.; Smith, A.; Sonnhammer, E.; Staden, R.; Sulston
proat, J.; Wohldman, P.
A;Description: The C. elegans genome project: Contiguous nucleotide sequence of over
A;Reference number: S72566
A;Accession: S72579
A;Molecule type: DNA
A;Residues: 1-412 <WIL>
A;Cross-references: EMBL:U21324; NID:g687879; PID:g687893
A;Experimental source: strain Bristol N2
C;Genetics:
A;Map position: 3
A;Introns: 58/3; 106/1; 150/1
A;Note: C35D10.14
C;Superfamily: C-type lectin homology
F;159-275/Domain: C-type lectin homology <LCH>

  Query Match             9.2%;  Score 78;  DB 2;  Length 412;
  Best Local Similarity   25.5%;  Pred. No. 15;
  Matches   41;  Conservative   23;  Mismatches   59;  Indels   38;  Gaps    9;

Qy      28 TDYVTLQRGSAYGG---VLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK----PWPV 80
           :||  : :   :||   :|||:  :|   |  : || |:  | :   |   |    | ||
Db     231 SDYTNIDQSFGHGGNCALLSNYQPSVVPMGYWFSFPCTAAVPFICKRSAGIKCSGTPPPV 290

Qy      81 ALYLTPVSSAGGLVIKAGSLIAVLILRQTN---NYNSDDF-----------------QF 119
              :||| |       :  |||  ::   |   ||::: |                  |
Db     291 T--VTPVPSNPSF-CNSTLLIAPGVITSPNYPSNYDNNQFCSYHLSTLGSYRVLLHFSAF 347

Qy     120 VWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTV 160
           :  | |:||   |   |:  :      | ||:  | :|
Db     348 ATEV--NFDLVVVYNGESASSSQI----GRYSGSLD-PFSV 381


RESULT 34
H41662
```

```
150K mating aggregate formation protein Asc10 precursor - Enterococcus faecalis plasmid
C;Species: Enterococcus faecalis
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 26-Aug-1999
C;Accession: H41662
R;Kao, S.M.; Olmsted, S.B.; Viksnins, A.S.; Gallo, J.C.; Dunny, G.M.
J. Bacteriol. 173, 7650-7664, 1991
A;Title: Molecular and genetic analysis of a region of plasmid pCF10 containing positive
terococcus faecalis.
A;Reference number: A41662; MUID:92041679; PMID:1938961
A;Accession: H41662
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1305 <KAO>
A;Cross-references: GB:M64978; NID:g150552; PIDN:AAA65848.1; PID:g150555
C;Genetics:
A;Genome: plasmid
C;Superfamily: aggregation protein asa1

  Query Match             9.2%;  Score 78;  DB 2;  Length 1305;
  Best Local Similarity   21.4%;  Pred. No. 64;
  Matches   43;  Conservative   24;  Mismatches   58;  Indels   76;  Gaps    8;

Qy      11 VDLSTQIFCHNDYPET-----------------ITDYVTLQ----RGS----------- 37
           |||:| :   :|| ||                  ||:  |:     :|:
Db     963 VDLATGVSFFDDYDETKVTPIKDLLRVKDSKGVDITNQFTISWDDAKGTVTISAKDPQAF 1022

Qy      38 --AYGG----------VLSNFSGTVKYSGSSYPFPTTSETPRVV-----YNSRTDKPWPV 80
             ||||          | :: || | |     |    :|  ||      | : |    |
Db    1023 ILAYGGQELRVTLPTKVKADVSGDVYNSAEQNTFGQRIKTNTVVNHIPKVNPKKD----V 1078

Qy      81 ALYLTPVSSAGGLVIKAG----------------SLIAVLILRQTNNYNSDDFQFVWNI 123
           : :    |  |  || |                 ::    :    :   | |   |::
Db    1079 VIKVGDKQSQNGATIKLGEKFFYEFTSSDIPAEYAGVVEEWSISDKLDVKHDKFSGQWSV 1138

Qy     124 YANNDVVVPTG-----GCDVS 139
           :||:: |:  |     | |:|
Db    1139 FANSNFVLADGTKVNKGDDIS 1159


RESULT 35
HNVZ4X
hemagglutinin precursor - vaccinia virus (strain Copenhagen)
N;Alternate names: A56R protein
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
C;Accession: D42523
R;Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 517-563, 1990
A;Title: Appendix to "The complete DNA sequence of vaccinia virus".
A;Reference number: A42501
A;Accession: D42523
A;Molecule type: DNA
A;Residues: 1-315 <GOE>
A;Cross-references: GB:M35027; NID:g335317; PIDN:AAA48191.1; PID:g335539
A;Experimental source: strain Copenhagen
R;Goebel, S.J.; Johnson, G.P.; Perkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.
Virology 179, 247-266, 1990
A;Title: The complete DNA sequence of vaccinia virus.
A;Reference number: A42531; MUID:91021027; PMID:2219722
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
C;Superfamily: vaccinia virus hemagglutinin; immunoglobulin homology
C;Keywords: glycoprotein; hemagglutinin; late protein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-315/Product: hemagglutinin #status predicted <HEG>
F;27-105/Domain: immunoglobulin homology <IMM>
F;280-302/Domain: transmembrane #status predicted <TMM>
F;303-315/Domain: intracellular #status predicted <INT>
F;37,69,112,161,254/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 315;
  Best Local Similarity   21.3%;  Pred. No. 12;
  Matches   39;  Conservative   22;  Mismatches   53;  Indels   69;  Gaps    8;

Qy       2 VVNVGQNLVVDLSTQIFCHN--------DY--------------PETITDYVTLQRGSAY 39
           :||      :|:      |:         ||              || ||| |
Db     127 IVNTDSESTIDIILSGSTHSPETSSEKPDYIDNSNCSSVFEIATPEPITDNV-------- 178

Qy      40 GGVLSNFSGTVKY-----------SGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
                 : : || |           || |    || |||  : :   |       :  | ||
Db     179 ----EDHTDTVTYTSDSINTVSATSGES----TTDETPEPITDKEEDHTVTDTVSYTTVS 230

Qy      89 SAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLP 148
           :: |:|               |    :||   ::: | :|| |  |     :    | ::
Db     231 TSSGIV--------------TTKSTTDDAD-LYDTYNDNDTVPST-----TVGSSTTSIS 270

Qy     149 DYR 151
           :|:
Db     271 NYK 273


RESULT 36
HMIVS2
hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)
C;Species: influenza A virus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C;Accession: A29971
R;Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A;Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in Chin
A;Reference number: A94370; MUID:88101364; PMID:3336940
A;Accession: A29971
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M19056; NID:g324208
A;Note: the sequence in GenBank entry FLAHAPA, release 106, (PID:g324209) differs fro
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 550;
  Best Local Similarity   24.7%;  Pred. No. 24;
  Matches   36;  Conservative   17;  Mismatches   58;  Indels   35;  Gaps    6;

Qy      23 YPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK 76
           ||  : || :|:   |  | |        ::|  :  |||          |   : || :
Db      98 YPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSS----ACKRGPASGFFSRLN- 152

Qy      77 PWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANNDV 129
             |               : |:||   || :   || | |   ::| ::          |
Db     153 -W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTNLY 195

Qy     130 VVPTGGCDVSARDVTVTLPDYRGSVP 155
           |  :|   || |    |:    || |
Db     196 VQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 37
HMIV77
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: A27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
```

```
A;Accession: A27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16737; NID:g324081; PIDN:AAA43143.1; PID:g324082
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 24;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:         |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPASGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |              : |:||   || :   || | |   ::| ::         |
Db     152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTDQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || |    |:   || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 38
HMIV89
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: D27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: D27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 24;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | ||  : || :|:   |  | |        ::|  :  ||:         |   : ||
Db      96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPASGFFSRL 151

Qy      75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
           :  |               : |:||   || :   || | |   ::| ::         |
Db     152 N--W---------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy     128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
             |  :|   || |    |:   || |
Db     194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 39
JQ1153
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1153
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1153
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00929; NID:g221279; PIDN:BAA00769.1; PID:g221280
A;Note: the authors translated the codon GGG for residue 218 as Glu
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             9.1%;  Score 77.5;  DB 2;  Length 550;
  Best Local Similarity   24.7%;  Pred. No. 24;
  Matches   36;  Conservative   17;  Mismatches   58;  Indels   35;  Gaps    6;

Qy      23 YPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK 76
           ||  : || :|:   |  | |        ::|  :  |||         |   : || :
Db      98 YPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSS----ACKRGPASGFFSRLN- 152

Qy      77 PWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANNDV 129
            |              : |:||   || :   || | |   ::| ::         |
Db     153 -W---------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTNLY 195

Qy     130 VVPTGGCDVSARDVTVTLPDYRGSVP 155
           |  :|   || |    |:   || |
Db     196 VQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 40
HMIVHA
hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C;Accession: A93231; A04051
R;Verhoeyen, M.; Fang, R.; Min Jou, W.; Devos, R.; Huylebroeck, D.; Saman, E.; Fiers,
Nature 286, 771-776, 1980
A;Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strain
A;Reference number: A93231; MUID:80254693; PMID:7402351
A;Accession: A93231
A;Molecule type: genomic RNA
A;Residues: 1-566 <VER>
A;Cross-references: GB:J02090; NID:g324131; PIDN:AAA43178.1; PID:g324132
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             9.1%;  Score 77.5;  DB 1;  Length 566;
  Best Local Similarity   25.7%;  Pred. No. 25;
  Matches   44;  Conservative   21;  Mismatches   67;  Indels   39;  Gaps    9;

Qy       1 PVVNVGQNLVVDL---STQIFCHNDYPETITDYVTLQRGSAYGGVLS------NFSGTVK 51
           |  :| ||   ||    :: |  | ||  : || :|:   |  | |         ::|  :
Db      90 PHCDVFQNETWDLFVERSKAF-SNCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQ 148

Qy      52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNN 111
```

```
           ||:         |   : || :  |               : |:||   || :   ||
Db    149 NGGSN----ACKRGPGSGFFSRLN--W--------------LTKSGSTYPVLNVTMPNN 187

Qy    112 YNSDDFQFVWNIY---ANND----VVVPTGGCDVSARDVTVTLPDYRGSVP 155
          | |   ::| |:    | :     |  :|   || |    |:    || |
Db    188 DNFDKL-YIWGIHHPSTNQEQTSLYVQASGRVTVSTRRSQQTIIPNIGSRP 237


RESULT 41
D90176
hypothetical protein SSO0334 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: D90176
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: D90176
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-433 <KUR>
A;Cross-references: GB:AE006641; NID:g13813479; PIDN:AAK40667.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO0334

  Query Match             9.1%;  Score 77;  DB 2;  Length 433;
  Best Local Similarity   22.7%;  Pred. No. 20;
  Matches   40;  Conservative   22;  Mismatches   42;  Indels   72;  Gaps   10;

Qy     24 PETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP---------------------- 60
          | | | :      : :||     |:||  | | || |
Db    188 PNTWTSW------TWFGG---TFNGTA-YIGGSYKFVDHWLYGNGVMVNVQFFKPVIWIF 237

Qy     61 TTSETPRVVYNSRTDKPWPVALY----LTPVS--------SAGGLVIKAGSLIAVLILRQ 108
          :||||     ::| ||  |::     ||||:        : ||: : ||::|
Db    238 ESSETP----VAQTPKPVQVSVSSVYGLTPVNVYSYEINGTQGGVAV-AGNMIV------ 286

Qy    109 TNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVP---IPLTVY 161
                      |  |  :| |:|   |    :    |   |: |   :|:  :
Db    287 ------------TIQPNTKIVTPSGNLSVFNFSIVFYNPANAGNAPNGQVPILAF 329


RESULT 42
T00467
probable UDPglucose 4-epimerase (EC 5.1.3.2) [imported] - Arabidopsis thaliana
N;Alternate names: UDPgalactose 4-epimerase homolog
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
C;Accession: T00467; F84761
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, April 1998
A;Description: Arabidopsis thaliana chromosome II BAC F19I3 genomic sequence.
A;Reference number: Z14160
A;Accession: T00467
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-375 <ROU>
A;Cross-references: EMBL:AC004238; NID:g3033373; PID:g3033381
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: F84761
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-375 <STO>
A;Cross-references: GB:AE002093; NID:g3033381; PIDN:AAC12825.1; GSPDB:GN00139
C;Genetics:
A;Gene: F19I3.8; At2g34850
A;Map position: 2
A;Introns: 29/3; 67/3; 103/3; 124/3; 203/2; 246/3; 295/1
C;Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homolo
C;Keywords: isomerase

  Query Match             9.0%;  Score 76.5;  DB 2;  Length 375;
  Best Local Similarity   28.0%;  Pred. No. 19;
  Matches   28;  Conservative   14;  Mismatches   43;  Indels   15;  Gaps    3;

Qy     18 FCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKP 77
          : ||    |:   | |:  :|:|     :| |    |     | | |||:|
Db    125 YYHNITSNTL---VVLETMAAHGVKTLIYSSTCATYGEPEKMPITEETPQV--------- 172

Qy     78 WPVALYLTPVSSAGGLVI--KAGSLIAVLILRQTNNYNSD 115
           |:  |      |  :::     |::||:|||  |   ||
Db    173 -PINPYGKAKKMAEDIILDFSKNSIMAVMILRYFNVIGSD 211


RESULT 43
JQ1154
hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragmen
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1154
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1154
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00930; NID:g221273; PIDN:BAA00770.1; PID:g221274
A;Note: the authors translated the codon GGG for residue 218 as Glu
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             9.0%;  Score 76.5;  DB 2;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 30;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;

Qy     21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
          | ||  : || :|:   |  | |        ::|  :  ||:          |   : ||
Db     96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPANGFFSRL 151

Qy     75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
          :  |               : |:||   || :   || | |   ::| ::         |
Db    152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy    128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
            |  :|   || |    |:    || |
Db    194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 44
JQ1155
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1155
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1155
A;Molecule type: genomic RNA
```

```
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00931; NID:g221277; PIDN:BAA00771.1; PID:g221278
A;Note: the authors translated the codon GGG for residue 218 as Glu, GCC for residue 538
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             9.0%;  Score 76.5;  DB 2;  Length 550;
  Best Local Similarity   24.3%;  Pred. No. 30;
  Matches   36;  Conservative   18;  Mismatches   59;  Indels   35;  Gaps    6;

Qy     21 NDYPETITDYVTLQRGSAYGGVLS------NFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
          | ||  : || :|:   |  | |        ::|  :  ||:        |   : ||
Db     96 NCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQNGGSN----ACKRGPANGFFSRL 151

Qy     75 DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY-------ANN 127
          :  |               : |:||   || :   || | |   ::| ::         |
Db    152 N--W--------------LTKSGSTYPVLNVTMPNNDNFDKL-YIWGVHHPSTNQEQTN 193

Qy    128 DVVVPTGGCDVSARDVTVTLPDYRGSVP 155
            |  :|   || |    |:    || |
Db    194 LYVQASGRVTVSTRRSQQTIIPNIGSRP 221


RESULT 45
HMIVH
hemagglutinin precursor - influenza A virus
C;Species: influenza A virus
C;Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999
C;Accession: A93705; A93233; A04051; A93231; A94441
R;Both, G.W.; Sleigh, M.J.
Nucleic Acids Res. 8, 2561-2575, 1980
A;Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza
A;Reference number: A93705; MUID:81053698; PMID:6253883
A;Accession: A93705
A;Molecule type: genomic RNA
A;Residues: 1-566 <BOT>
A;Cross-references: GB:V01103
A;Experimental source: strain A/NT/60/68/29C
A;Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/6
R;Dopheide, T.A.; Ward, C.W.
FEBS Lett. 110, 181-183, 1980
A;Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.
A;Reference number: A91276; MUID:80179105; PMID:6768586
A;Contents: annotation; disulfide bonds
R;Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.
Nature 287, 301-306, 1980
A;Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from
A;Reference number: A93233; MUID:81030852; PMID:7421990
A;Accession: A93233
A;Molecule type: genomic RNA
A;Residues: 1-24,'S',26,'D',28-159,'G',161-197,'I',199-241,'L',243-249 <GET>
A;Experimental source: strain X-31[H3]
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;536-552/Domain: transmembrane #status predicted <TM1>
F;30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             9.0%;  Score 76.5;  DB 1;  Length 566;
  Best Local Similarity   25.1%;  Pred. No. 31;
  Matches   43;  Conservative   22;  Mismatches   67;  Indels   39;  Gaps    9;

Qy      1 PVVNVGQNLVVDL---STQIFCHNDYPETITDYVTLQRGSAYGGVLS------NFSGTVK 51
          |  :| ||   ||    :: |  | ||  : || :|:   |  | |        ::|  :
Db     90 PHCDVFQNETWDLFVERSKAF-SNCYPYDVPDYASLRSLVASSGTLEFITEGFTWTGVTQ 148

Qy     52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNN 111
            ||:         |   : || :  |               : |:||   || :   ||
Db    149 NGGSN----ACKRGPDSGFFSRLN--W--------------LTKSGSTYPVLNVTMPNN 187

Qy    112 YNSDDFQFVWNIY---ANND----VVVPTGGCDVSARDVTVTLPDYRGSVP 155
            | |   ::| ::    | :     |  :|   || |    |:    || |
Db    188 DNFDKL-YIWGVHHPSTNQEQTSLYVQASGRVTVSTRRSQQTIIPNIGSRP 237


Search completed: November 28, 2002, 18:53:57
Job time : 47 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 17:18:40 ; Search time 25 Seconds
(without alignments)
189.484 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	96.5	11.4	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
2	83	9.8	499	2	US-09-032-315-3	Sequence 3, Appli
3	83	9.8	499	2	US-08-993-318A-3	Sequence 3, Appli
4	83	9.8	499	4	US-09-399-886-3	Sequence 3, Appli
5	83	9.8	499	4	US-09-396-260-3	Sequence 3, Appli
6	83	9.8	499	4	US-09-576-281-3	Sequence 3, Appli
7	83	9.8	519	1	US-08-462-484-4	Sequence 4, Appli
8	83	9.8	519	1	US-08-441-147-4	Sequence 4, Appli
9	83	9.8	519	5	PCT-US95-07536-4	Sequence 4, Appli
10	79	9.3	373	1	US-07-723-002C-2	Sequence 2, Appli
11	78	9.2	178	4	US-09-134-001C-4994	Sequence 4994, Ap
12	75	8.8	1612	1	US-08-169-927-2	Sequence 2, Appli
13	73	8.6	411	1	US-07-937-609-21	Sequence 21, Appl
14	73	8.6	411	4	US-08-029-170-21	Sequence 21, Appl
15	73	8.6	451	4	US-08-430-286A-10	Sequence 10, Appl
16	72	8.5	1112	2	US-08-714-402-2	Sequence 2, Appli
17	72	8.5	1161	4	US-09-327-536-2	Sequence 2, Appli
18	71	8.4	261	4	US-09-110-959A-2	Sequence 2, Appli
19	70	8.3	552	3	US-09-295-186-10	Sequence 10, Appl
20	70	8.3	573	3	US-09-295-186-11	Sequence 11, Appl
21	69.5	8.2	422	1	US-08-190-802A-53	Sequence 53, Appl
22	69.5	8.2	422	1	US-08-190-802A-55	Sequence 55, Appl
23	69.5	8.2	422	4	US-08-477-346-53	Sequence 53, Appl
24	69.5	8.2	422	4	US-08-477-346-55	Sequence 55, Appl
25	69.5	8.2	422	4	US-08-473-089-53	Sequence 53, Appl
26	69.5	8.2	422	4	US-08-473-089-55	Sequence 55, Appl
27	69.5	8.2	422	4	US-08-487-072A-53	Sequence 53, Appl
28	69.5	8.2	422	4	US-08-487-072A-55	Sequence 55, Appl
29	69	8.1	332	4	US-09-331-581-23	Sequence 23, Appl
30	69	8.1	379	1	US-07-723-002C-4	Sequence 4, Appli
31	69	8.1	422	2	US-08-712-072C-2	Sequence 2, Appli
32	69	8.1	619	4	US-09-620-412C-309	Sequence 309, App
33	68.5	8.1	129	4	US-09-071-035-198	Sequence 198, App
34	68.5	8.1	246	4	US-09-216-295-20	Sequence 20, Appl
35	68.5	8.1	415	4	US-09-032-523-2	Sequence 2, Appli
36	68.5	8.1	686	3	US-08-947-965-73	Sequence 73, Appl
37	68	8.0	291	6	5194600-2	Patent No. 5194600
38	68	8.0	624	4	US-09-071-035-244	Sequence 244, App
39	68	8.0	725	2	US-08-816-105A-1	Sequence 1, Appli
40	68	8.0	1223	4	US-09-071-035-236	Sequence 236, App
41	68	8.0	1301	4	US-09-071-035-234	Sequence 234, App
42	68	8.0	1301	4	US-09-071-035-238	Sequence 238, App
43	68	8.0	1301	4	US-09-071-035-242	Sequence 242, App
44	68	8.0	1349	2	US-08-612-734B-2	Sequence 2, Appli
45	68	8.0	3739	3	US-09-320-878-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 11.4%; Score 96.5; DB 4; Length 10182;
Best Local Similarity 27.2%; Pred. No. 2.1;
Matches 49; Conservative 32; Mismatches 62; Indels 37; Gaps 11;

```
Qy       2 VVNVGQNLVVDL---------STQIFCHNDYPETITDYVTLQRG--SAYGGVLSNFSGTV 50
           | || ||  | |          :|     ::   |:|   ||  |  :|   : :| : |
Db    1924 VRNVPQNAQVTLYQSNGTVIPNTNTTIDSNGIATVTIQGTLPTGNITAKTSMTNNVTYT- 1982

Qy      51 KYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTN 110
           | : |     || :    |::  :|:          |:   |: |     : |:: |
Db    1983 KQNSSGIASNTTEDIS--VFSENSDQ----------VNVTAGMQAKND---GIKIIKGT- 2026

Qy     111 NYNSDDF-QFVWNIYANNDVV---VPT---GGCDVSARDVTVTLPDYRG--SVPIPLTVY 161
           ||| :||  |: || |:: :     |         : : ||||||:::|  :| ||:|:|
Db    2027 NYNFNDFNSFISNIPAHSTLTWNEEPNSWKNNIGTTTKTVTVTLPNHQGTRTVDIPITIY 2086
```

RESULT 2
US-09-032-315-3
; Sequence 3, Application US/09032315
; Patent No. 5985818
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Xu, Feng
; TITLE OF INVENTION: LACCASE MUTANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5985818o No. 5985818disk of No. 5985818th America, Inc.

```
;       STREET:  405 Lexington Avenue
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10174
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/032,315
;       FILING DATE:  27-FEB-1998
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Rozek, Carol
;       REGISTRATION NUMBER:  36,993
;       REFERENCE/DOCKET NUMBER:  5200.200-US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-867-0123
;       TELEFAX:  212-878-9655
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  499 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-032-315-3

  Query Match             9.8%;  Score 83;  DB 2;  Length 499;
  Best Local Similarity   28.5%;  Pred. No. 0.82;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP---ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db     229 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 284

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||        |    |:|
Db     285 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 335

Qy     116 DFQFVWNIYANNDVVVP 132
           :|      : ||    |
Db     336 NF------FINNATFTP 346


RESULT 3
US-08-993-318A-3
; Sequence 3, Application US/08993318A
; Patent No. 5998353
;  GENERAL INFORMATION:
;    APPLICANT:  Pedersen, Anders
;    APPLICANT:  Svendsen, Allan
;    APPLICANT:  Schneider, Palle
;    APPLICANT:  Rasmussen, Grethe
;    APPLICANT:  Cherry, Joel
;    TITLE OF INVENTION:  LACCASE MUTANTS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5998353o No. 5998353disk of No. 5998353th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/993,318A
;      FILING DATE:  December 18, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gregg, Valeta A.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  5032.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-993-318A-3

  Query Match             9.8%;  Score 83;  DB 2;  Length 499;
  Best Local Similarity   28.5%;  Pred. No. 0.82;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db     229 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 284

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||        |    |:|
Db     285 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 335

Qy     116 DFQFVWNIYANNDVVVP 132
           :|      : ||    |
Db     336 NF------FINNATFTP 346


RESULT 4
US-09-399-886-3
; Sequence 3, Application US/09399886
; Patent No. 6140092
;  GENERAL INFORMATION:
;    APPLICANT:  Pedersen, Anders
;    APPLICANT:  Svendsen, Allan
;    APPLICANT:  Schneider, Palle
;    APPLICANT:  Rasmussen, Grethe
;    APPLICANT:  Cherry, Joel
;    TITLE OF INVENTION:  LACCASE MUTANTS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6140092o No. 6140092disk of No. 6140092th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/399,886
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/993,318
;      FILING DATE:  December 18, 1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gregg, Valeta A.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  5032.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;      TELEX:
```

```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-399-886-3

  Query Match             9.8%;  Score 83;  DB 4;  Length 499;
  Best Local Similarity   28.5%;  Pred. No. 0.82;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db     229 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 284

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||        |    |:|
Db     285 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 335

Qy     116 DFQFVWNIYANNDVVVP 132
          :|      : ||    |
Db     336 NF------FINNATFTP 346


RESULT 5
US-09-396-260-3
; Sequence 3, Application US/09396260
; Patent No. 6184015
;  GENERAL INFORMATION:
;    APPLICANT:  Svendsen, Allan
;    APPLICANT:  Xu, Feng
;    TITLE OF INVENTION:  LACCASE MUTANTS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6184015o No. 6184015disk of No. 6184015th America, Inc.
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/396,260
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/032,315
;      FILING DATE:  27-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rozek, Carol
;      REGISTRATION NUMBER:  36,993
;      REFERENCE/DOCKET NUMBER:  5200.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-396-260-3

  Query Match             9.8%;  Score 83;  DB 4;  Length 499;
  Best Local Similarity   28.5%;  Pred. No. 0.82;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db     229 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 284

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||        |    |:|
Db     285 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 335

Qy     116 DFQFVWNIYANNDVVVP 132
          :|      : ||    |
Db     336 NF------FINNATFTP 346


RESULT 6
US-09-576-281-3
; Sequence 3, Application US/09576281
; Patent No. 6277611
;  GENERAL INFORMATION:
;    APPLICANT:  Pedersen, Anders
;    APPLICANT:  Svendsen, Allan
;    APPLICANT:  Schneider, Palle
;    APPLICANT:  Rasmussen, Grethe
;    APPLICANT:  Cherry, Joel
;    TITLE OF INVENTION:  LACCASE MUTANTS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6277611o No. 6277611disk of No. 6277611th America
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/576,281
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/993,318
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gregg, Valeta A.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  5032.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-576-281-3

  Query Match             9.8%;  Score 83;  DB 4;  Length 499;
  Best Local Similarity   28.5%;  Pred. No. 0.82;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db     229 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 284

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||        |    |:|
Db     285 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 335
```

```
Qy    116 DFQFVWNIYANNDVVVP 132
          :|      : ||    |
Db    336 NF------FINNATFTP 346


RESULT 7
US-08-462-484-4
; Sequence 4, Application US/08462484
; Patent No. 5667531
;  GENERAL INFORMATION:
;    APPLICANT:  Yaver, Debbie Sue
;    APPLICANT:  Xu, Feng
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Schneider, Palle
;    APPLICANT:  Aaslyng, Dorrit A.
;    TITLE OF INVENTION:  PURIFIED POLYPORUS LACCASES AND
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5667531o No. 5667531disk of No. 5667531th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,484
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/441,147
;      FILING DATE:  15-MAY-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney, Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  4185.010-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 878 9655
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  519 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-462-484-4

  Query Match             9.8%;  Score 83;  DB 1;  Length 519;
  Best Local Similarity   28.5%;  Pred. No. 0.87;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy      7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
          | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db    249 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 304

Qy     56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||        |    |:|
Db    305 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 355

Qy    116 DFQFVWNIYANNDVVVP 132
          :|      : ||    |
Db    356 NF------FINNATFTP 366


RESULT 8
US-08-441-147-4
; Sequence 4, Application US/08441147
; Patent No. 5770418
;  GENERAL INFORMATION:
;    APPLICANT:  Yaver, Debbie Sue
;    APPLICANT:  Xu, Feng
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Schneider, Palle
;    APPLICANT:  Aaslyng, Dorrit A.
;    TITLE OF INVENTION:  PURIFIED POLYPORUS LACCASES AND
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5770418o No. 5770418disk of No. 5770418th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,147
;      FILING DATE:  15-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney, Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  4185.010-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 878 9655
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  519 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-441-147-4

  Query Match             9.8%;  Score 83;  DB 1;  Length 519;
  Best Local Similarity   28.5%;  Pred. No. 0.87;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy      7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
          | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db    249 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 304

Qy     56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||        |    |:|
Db    305 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 355

Qy    116 DFQFVWNIYANNDVVVP 132
          :|      : ||    |
Db    356 NF------FINNATFTP 366


RESULT 9
PCT-US95-07536-4
; Sequence 4, Application PC/TUS9507536
;  GENERAL INFORMATION:
;    APPLICANT:
;    APPLICANT:
;    TITLE OF INVENTION:  PURIFIED POLYPORUS LACCASES AND
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Novo Nordisk of North America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07536
;      FILING DATE:  15-June-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/265,534
;      FILING DATE:  24-June-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney, Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  4185.204-WO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 878 9655
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  519 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-07536-4

  Query Match             9.8%;  Score 83;  DB 5;  Length 519;
  Best Local Similarity   28.5%;  Pred. No. 0.87;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db     249 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 304

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||        |    |:|
Db     305 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 355

Qy     116 DFQFVWNIYANNDVVVP 132
           :|      : ||    |
Db     356 NF------FINNATFTP 366


RESULT 10
US-07-723-002C-2
; Sequence 2, Application US/07723002C
; Patent No. 5447862
;  GENERAL INFORMATION:
;    APPLICANT:  Heim, Jutta
;    APPLICANT:  Meyhack, Bernd
;    APPLICANT:  Gysler, Christof
;    APPLICANT:  Visser, Jacob
;    APPLICANT:  Kester, Hermanus Cornelis Maria
;    TITLE OF INVENTION:  No. 5447862el Expression System
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/723,002C
;      FILING DATE:  28-JUN-1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8702475
;      FILING DATE:  04-FEB-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 150,880
;      FILING DATE:  29-JAN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8818046.8
;      FILING DATE:  28-JUL-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8914666.6
;      FILING DATE:  26-JUN-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 384,898
;      FILING DATE:  24-JUL-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Foley, Shawn P.
;      REGISTRATION NUMBER:  33,071
;      REFERENCE/DOCKET NUMBER:  4-16317/+/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8615
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  373 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-723-002C-2

  Query Match             9.3%;  Score 79;  DB 1;  Length 373;
  Best Local Similarity   23.6%;  Pred. No. 1.5;
  Matches   38;  Conservative   25;  Mismatches   72;  Indels   26;  Gaps    7;

Qy       9 LVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGV-LSNFSGTVKYSGSSYPFPTTSETPR 67
           |  |  :::   |:|    :||     |  |  | |   |  | :|| :| : |:   |:
Db     200 LGTDADSRVSITNNYINGESDYSATCDGHHYWNVYLDGSSDKVTFSG-NYLYKTSGRAPK 258

Qy      68 V-------VYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIA-VLILRQTNN-----YNS 114
           |       :||:  :      |  :     :|| |:  |:  : |  : :|:      ::|
Db     259 VQDNTYLHIYNNYWENNSGHAFEI----GSGGYVLAEGNYFSNVDTVLETDTFEGALFSS 314

Qy     115 DDFQFVWNIYANNDVVVPTGGCDVSARDVTV-------TLP 148
           |       |     |    | |::    ||       |||
Db     315 DSASSTCESYIGRSCVANVNGGDLTGTSTTVLSNLSGDTLP 355


RESULT 11
US-09-134-001C-4994
; Sequence 4994, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4994
;   LENGTH: 178
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4994

  Query Match             9.2%;  Score 78;  DB 4;  Length 178;
  Best Local Similarity   31.2%;  Pred. No. 0.66;
  Matches   20;  Conservative   11;  Mismatches   19;  Indels   14;  Gaps    2;

Qy      96 KAGSLIAVLILRQTNNYNSDDFQFV------------WNIYANN-DVVVPTGGCDVSAR 141
           | | ||  || :  |: : : :: |             |   :|  ||:: |||  :| |
Db      34 KGGQLIQSLINQHENHVDINHYRIVKDDIEAITHQLKLWLTSSNQLDVIITTGGTGISQR 93
```

```
Qy     142 DVTV 145
           |||:
Db      94 DVTI 97


RESULT 12
US-08-169-927-2
; Sequence 2, Application US/08169927
; Patent No. 5783441
;  GENERAL INFORMATION:
;    APPLICANT:  Carl, Mitchell
;    APPLICANT:  Dobson, Michael E.
;    APPLICANT:  Ching, Wei Mei
;    APPLICANT:  Dasch, Gregory A
;    TITLE OF INVENTION:  Gene and Protein Applicable to the
;    TITLE OF INVENTION:  Preparation of Vaccines for Rickettsia prowazekii and
;    TITLE OF INVENTION:  Rickettsia typhi and the Detection of Both
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Counsel, Naval Medical R & D Command
;      STREET:  Bldg. 1, T-12, 8901 Wisconsin Ave.
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20889-5606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/169,927
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/742,128
;      FILING DATE:  08/09/91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, A. David
;      REGISTRATION NUMBER:  24,743
;      REFERENCE/DOCKET NUMBER:  75,976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-1022
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1612 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-169-927-2

  Query Match             8.8%;  Score 75;  DB 1;  Length 1612;
  Best Local Similarity   23.7%;  Pred. No. 34;
  Matches   31;  Conservative   17;  Mismatches   47;  Indels   36;  Gaps    5;

Qy      31 VTLQRGSAYGGV--LSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVS 88
           :|:   : | |:    |  |||  ||          |  :|
Db     908 ITVNANTLYSGIRTTKNNQGTVTLSGG------MPNNPGTIY----------------- 943

Qy      89 SAGGLVIKAGSLIAVLILRQ---TNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTV 145
             || :: ||       |:|   | :||:       |:  |:|| : |||   :  |  :
Db     944 ---GLGLENGS----PKLKQVTFTTDYNNLGSIIANNVTINDDVTLTTGGIAGTDFDAKI 996

Qy     146 TLPDYRGSVPI 156
           ||    |:  :
Db     997 TLGSVNGNANV 1007


RESULT 13
US-07-937-609-21
; Sequence 21, Application US/07937609
; Patent No. 5319073
;  GENERAL INFORMATION:
;    APPLICANT:  WANK, Stephen A.
;    TITLE OF INVENTION:  CLONING AND FUNCTIONAL EXPRESSION OF
;    TITLE OF INVENTION:  CHOLECYSTOKININ RECEPTOR-ENCODING DNA
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/937,609
;      FILING DATE:  19920902
;      CLASSIFICATION:  436
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/831,248
;      FILING DATE:  07-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/861,769
;      FILING DATE:  01-APR-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/928,033
;      FILING DATE:  11-AUG-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  40399/166 NIHD
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  411 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-937-609-21

  Query Match             8.6%;  Score 73;  DB 1;  Length 411;
  Best Local Similarity   28.3%;  Pred. No. 7.8;
  Matches   28;  Conservative   16;  Mismatches   23;  Indels   32;  Gaps    6;

Qy      27 ITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK----PWPVAL 82
           :|::: || |       |||      |:   | |:: |  |  : |::     | :||
Db      30 VTEWLALQAG--------NFS-------SALGLPATTQAPSQVRANLTNQFVQPSWRIAL 74

Qy      83 YLTPVSSAGGLVIKA---GSLIAVLI------LRQTNNY 112
           :    | | |||:     |:|| : |      :|   ||
Db      75 W----SLAYGLVVAVAVFGNLIVIWIILAHKRMRTVTNY 109


RESULT 14
US-08-029-170-21
; Sequence 21, Application US/08029170
; Patent No. 6169173
;  GENERAL INFORMATION:
;    APPLICANT:  WANK, Stephen A.
;    TITLE OF INVENTION:  CLONING AND FUNCTIONAL EXPRESSION OF
;    TITLE OF INVENTION:  CHOLECYSTOKININ RECEPTOR-ENCODING DNA
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
```

```
;       ZIP:  22313-0299
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/029,170
;       FILING DATE:  19930310
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/937,609
;       FILING DATE:  02-SEP-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/928,033
;       FILING DATE:  11-AUG-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/861,769
;       FILING DATE:  01-APR-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/831,248
;       FILING DATE:  07-FEB-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BENT, Stephen A.
;       REGISTRATION NUMBER:  29,768
;       REFERENCE/DOCKET NUMBER:  40399/166 NIHD
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703)836-9300
;       TELEFAX:  (703)683-4109
;       TELEX:  899149
;   INFORMATION FOR SEQ ID NO:  21:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  411 amino acids
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
US-08-029-170-21

  Query Match             8.6%;  Score 73;  DB 4;  Length 411;
  Best Local Similarity   28.3%;  Pred. No. 7.8;
  Matches   28;  Conservative   16;  Mismatches   23;  Indels   32;  Gaps    6;

Qy      27 ITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK----PWPVAL 82
           :|::: || |        |||       |:   | |:: |  |  : |::     | :||
Db      30 VTEWLALQAG--------NFS-------SALGLPATTQAPSQVRANLTNQFVQPSWRIAL 74

Qy      83 YLTPVSSAGGLVIKA---GSLIAVLI------LRQTNNY 112
           :    | | |||:     |:|| : |      :|   ||
Db      75 W----SLAYGLVVAVAVFGNLIVIWIILAHKRMRTVTNY 109


RESULT 15
US-08-430-286A-10
; Sequence 10, Application US/08430286A
; Patent No. 6225080
;  GENERAL INFORMATION:
;    APPLICANT:  Uhl, George R.
;    APPLICANT:  Eppler, C. Mark
;    APPLICANT:  Wang, Jai-Bel
;    TITLE OF INVENTION:  Mu-Subtype Opioid Receptor
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby & Darby PC
;      STREET:  805 Third Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  US
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,286A
;      FILING DATE:  28-APR-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Robinson, Joseph R.
;      REGISTRATION NUMBER:  33,448
;      REFERENCE/DOCKET NUMBER:  0646/1A843-US5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-527-7700
;      TELEFAX:  212-753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  451 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Rattus rattus
;    IMMEDIATE SOURCE:
;      CLONE:  NEU-K
US-08-430-286A-10

  Query Match             8.6%;  Score 73;  DB 4;  Length 451;
  Best Local Similarity   28.3%;  Pred. No. 9;
  Matches   28;  Conservative   16;  Mismatches   23;  Indels   32;  Gaps    6;

Qy      27 ITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDK----PWPVAL 82
           :|::: || |        |||       |:   | |:: |  |  : |::     | :||
Db      30 VTEWLALQAG--------NFS-------SALGLPATTQAPSQVRANLTNQFVQPSWRIAL 74

Qy      83 YLTPVSSAGGLVIKA---GSLIAVLI------LRQTNNY 112
           :    | | |||:     |:|| : |      :|   ||
Db      75 W----SLAYGLVVAVAVFGNLIVIWIILAHKRMRTVTNY 109


RESULT 16
US-08-714-402-2
; Sequence 2, Application US/08714402
; Patent No. 5910441
;  GENERAL INFORMATION:
;    APPLICANT:  ROCHA, Claudia
;    APPLICANT:  FISCHETTI, Vincent A.
;    TITLE OF INVENTION:  FIBRONECTIN AND FIBRINOGEN BINDING
;    TITLE OF INVENTION:  PROTEIN FROM GROUP A STREPTOCOCCI
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,402
;      FILING DATE:  16-SEP-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  016921-097
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  1112 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-714-402-2

  Query Match             8.5%;  Score 72;  DB 2;  Length 1112;
  Best Local Similarity   22.5%;  Pred. No. 43;
  Matches   29;  Conservative   19;  Mismatches   31;  Indels   50;  Gaps    7;

Qy      26 TITDYV------------------TLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSET 65
           | |||:                    |:  |     :|||  |:     :|      |
Db     249 TFTDYIAGLDKVQLSAELSLFLENKEVLENTS-----ISNFKSTIGGQEITY-----KGT 298

Qy      66 PRVVYNSRTDKPWPVALYLT-PVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY 124
           |:| : : |    : |:|  :|: || :               :||::  :||| :|
Db     299 VNVLYGNESTKE---SNYITNGLSNVGGSI--------------ESYNTETGEFVWYVY 340

Qy     125 AN-NDVVVP 132
           | |   :|
Db     341 VNPNRTNIP 349


RESULT 17
US-09-327-536-2
; Sequence 2, Application US/09327536
; Patent No. 6355477
; GENERAL INFORMATION:
;  APPLICANT:  FISCHETTI, Vincent A.
;  APPLICANT:  ROCHA, Claudia
;  TITLE OF INVENTION:  FIBRONECTIN AND FIBRINOGEN BINDING PROTEIN FROM GROUP A
;  TITLE OF INVENTION:  STREPTOCOCCI
;  FILE REFERENCE:  022927-008
;  CURRENT APPLICATION NUMBER: US/09/327,536
;  CURRENT FILING DATE:  1999-06-08
;  PRIOR APPLICATION NUMBER:  US 08/714,402
;  PRIOR FILING DATE:  1996-09-16
;  NUMBER OF SEQ ID NOS:  2
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1161
;   TYPE: PRT
;   ORGANISM: SFFBP gene
US-09-327-536-2

  Query Match             8.5%;  Score 72;  DB 4;  Length 1161;
  Best Local Similarity   22.5%;  Pred. No. 45;
  Matches   29;  Conservative   19;  Mismatches   31;  Indels   50;  Gaps    7;

Qy      26 TITDYV------------------TLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSET 65
           | |||:                    |:  |     :|||  |:     :|      |
Db     249 TFTDYIAGLDKVQLSAELSLFLENKEVLENTS-----ISNFKSTIGGQEITY-----KGT 298

Qy      66 PRVVYNSRTDKPWPVALYLT-PVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIY 124
           |:| : : |    : |:|  :|: || :               :||::  :||| :|
Db     299 VNVLYGNESTKE---SNYITNGLSNVGGSI--------------ESYNTETGEFVWYVY 340

Qy     125 AN-NDVVVP 132
           | |   :|
Db     341 VNPNRTNIP 349


RESULT 18
US-09-110-959A-2
; Sequence 2, Application US/09110959A
; Patent No. 6268197
; GENERAL INFORMATION:
;  APPLICANT:  Schulein, Martin
;  APPLICANT:  Outtrup, Helle
;  APPLICANT:  Jorgensen, Per Lina
;  APPLICANT:  Bjornvad, Mads Eskelund
;  TITLE OF INVENTION:  Alkaline Xyloglucanase
;  FILE REFERENCE:  5206.200-US
;  CURRENT APPLICATION NUMBER: US/09/110,959A
;  CURRENT FILING DATE:  1998-07-07
;  PRIOR APPLICATION NUMBER:  0822/97
;  PRIOR FILING DATE:  1997-07-07
;  PRIOR APPLICATION NUMBER:  1213/97
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER:  60/054,039
;  PRIOR FILING DATE:  1997-07-28
;  PRIOR APPLICATION NUMBER:  60/063,694
;  PRIOR FILING DATE:  1997-10-28
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 261
;   TYPE: PRT
;   ORGANISM: Bacillus licheniformis ATCC 14580
US-09-110-959A-2

  Query Match             8.4%;  Score 71;  DB 4;  Length 261;
  Best Local Similarity   21.8%;  Pred. No. 6.8;
  Matches   34;  Conservative   23;  Mismatches   65;  Indels   34;  Gaps    5;

Qy       4 NVGQNLVVDLSTQIFCHN----DYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPF 59
           ::  |   : :  |: ||     :    || : :   :   |   ::  ||   | |:
Db     125 SISANGTYNAAYDIWLHNTNKASWDSAPTDEIMIWLNNTNAGPAGSYVETVSIGGHSWK- 183

Qy      60 PTTSETPRVVYNSRTD----KPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
                ||    |    | | |  ::            | :  | | :|   || :|
Db     184 ---------VYKGYIDAGGGKGWNVFSFIR----------TANTQSANLNIRDFTNYLAD 224

Qy     116 DFQ------FVWNIYANNDVVVPTGGCDVSARDVTV 145
             |      :| ::    :|   ||  ::|  ||||
Db     225 SKQWLSKTKYVSSVEFGTEVFGGTGQINISNWDVTV 260


RESULT 19
US-09-295-186-10
; Sequence 10, Application US/09295186B
; Patent No. 6127137
; GENERAL INFORMATION:
;  APPLICANT:  Hasida, Miyoko
;  APPLICANT:  Tsutsumi, No. 6127137iko
;  APPLICANT:  Halkier, Torben
;  APPLICANT:  Stringer, Mary Ann
;  TITLE OF INVENTION:  An Acidic Phospholipase, Production, and
;  TITLE OF INVENTION:  Methods of Using Thereof (As Amended)
;  FILE REFERENCE:  4953.204-US
;  CURRENT APPLICATION NUMBER: US/09/295,186B
;  CURRENT FILING DATE:  1999-04-20
;  PRIOR APPLICATION NUMBER:  1215/96
;  PRIOR FILING DATE:  1996-10-31
;  PRIOR APPLICATION NUMBER:  PCT/DK97/00490
;  PRIOR FILING DATE:  1997-10-30
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 10
;   LENGTH: 552
;   TYPE: PRT
;   ORGANISM: Hyphozyma sp. CBS 648.91
US-09-295-186-10

  Query Match             8.3%;  Score 70;  DB 3;  Length 552;
  Best Local Similarity   23.9%;  Pred. No. 26;
  Matches   33;  Conservative   16;  Mismatches   45;  Indels   44;  Gaps    7;

Qy      23 YPETITDYVTLQ------RGSAYGGVL----SNFSGTVKYSGSSYPFPTTSETPRVVYNS 72
           |  ::|||  |        || ||       |:   | |:: :| |||       :: :
Db     265 YQTSLTDYFGLSIADKILNGSMYGNKFSVEWSDVKNTSKFTDASMPFPI------IIADE 318

Qy      73 RTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQF-VWNIYANNDVVV 131
           |                      : | ||   | | |  :  : ::|  ||   |    :
Db     319 R----------------------EPGELI---IPRNTTIWEFNPYEFGSWN--PNVSAFI 351
```

```
Qy    132 PTGGCDVSARDVTVTLPD 149
          |      |  : |  |||
Db    352 PIEILGSSLDNGTSVLPD 369


RESULT 20
US-09-295-186-11
; Sequence 11, Application US/09295186B
; Patent No. 6127137
; GENERAL INFORMATION:
;  APPLICANT:  Hasida, Miyoko
;  APPLICANT:  Tsutsumi, No. 6127137iko
;  APPLICANT:  Halkier, Torben
;  APPLICANT:  Stringer, Mary Ann
;  TITLE OF INVENTION:  An Acidic Phospholipase, Production, and
;  TITLE OF INVENTION:  Methods of Using Thereof (As Amended)
;  FILE REFERENCE:  4953.204-US
;  CURRENT APPLICATION NUMBER: US/09/295,186B
;  CURRENT FILING DATE:  1999-04-20
;  PRIOR APPLICATION NUMBER:  1215/96
;  PRIOR FILING DATE:  1996-10-31
;  PRIOR APPLICATION NUMBER:  PCT/DK97/00490
;  PRIOR FILING DATE:  1997-10-30
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 11
;   LENGTH: 573
;   TYPE: PRT
;   ORGANISM: Hyphozyma sp.  CBS 648.91
US-09-295-186-11

  Query Match             8.3%;  Score 70;  DB 3;  Length 573;
  Best Local Similarity   23.9%;  Pred. No. 27;
  Matches   33;  Conservative   16;  Mismatches   45;  Indels   44;  Gaps    7;

Qy     23 YPETITDYVTLQ------RGSAYGGVL----SNFSGTVKYSGSSYPFPTTSETPRVVYNS 72
          |  ::|||  |        || ||       |:   | |:: :| |||       :: :
Db    265 YQTSLTDYFGLSIADKILNGSMYGNKFSVEWSDVKNTSKFTDASMPFPI------IIADE 318

Qy     73 RTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQF-VWNIYANNDVVV 131
          |                    : | ||   | | |  :  : ::|  ||   |    :
Db    319 R---------------------EPGELI---IPRNTTIWEFNPYEFGSWN--PNVSAFI 351

Qy    132 PTGGCDVSARDVTVTLPD 149
          |      |  : |  |||
Db    352 PIEILGSSLDNGTSVLPD 369


RESULT 21
US-08-190-802A-53
; Sequence 53, Application US/08190802A
; Patent No. 5519003
;  GENERAL INFORMATION:
;    APPLICANT:  Mochly-Rosen, Daria
;    APPLICANT:  Ron, Dorit
;    TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  265
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  P.O. Box 60850
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-0850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/190,802A
;      FILING DATE:  01-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fabian, Gary R.
;      REGISTRATION NUMBER:  33,875
;      REFERENCE/DOCKET NUMBER:  8600-0139
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  MSL1, Fig. 36
US-08-190-802A-53

  Query Match             8.2%;  Score 69.5;  DB 1;  Length 422;
  Best Local Similarity   28.0%;  Pred. No. 20;
  Matches   26;  Conservative   18;  Mismatches   24;  Indels   25;  Gaps    6;

Qy     59 FP---TTSETPRVVYNSRT--DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYN 113
          ||   |||:  |:: :| |   ||    :|:: :|: |   ||  ||         ||::
Db     52 FPDLDTTSDEHRILLSSFTSSQKPEDETIYISKISTLGH--IKWSSL---------NNFD 100

Qy    114 SDDFQF-------VWNIYANNDVVV--PTGGCD 137
            |: :|       : :  ||: :  | | |:
Db    101 MDEMEFKPENSTRFPSKHLVNDISIFFPNGECN 133


RESULT 22
US-08-190-802A-55
; Sequence 55, Application US/08190802A
; Patent No. 5519003
;  GENERAL INFORMATION:
;    APPLICANT:  Mochly-Rosen, Daria
;    APPLICANT:  Ron, Dorit
;    TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  265
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  P.O. Box 60850
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-0850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/190,802A
;      FILING DATE:  01-FEB-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fabian, Gary R.
;      REGISTRATION NUMBER:  33,875
;      REFERENCE/DOCKET NUMBER:  8600-0139
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
```

```
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  ORF RB1, Fig. 38
US-08-190-802A-55

  Query Match             8.2%;  Score 69.5;  DB 1;  Length 422;
  Best Local Similarity   28.0%;  Pred. No. 20;
  Matches   26;  Conservative   18;  Mismatches   24;  Indels   25;  Gaps    6;

Qy      59 FP---TTSETPRVVYNSRT--DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYN 113
           ||   |||:  |:: :| |   ||    :|:: :|: |   ||  ||         ||::
Db      52 FPDLDTTSDEHRILLSSFTSSQKPEDETIYISKISTLGH--IKWSSL---------NNFD 100

Qy     114 SDDFQF-------VWNIYANNDVVV--PTGGCD 137
           |: :|         : :  ||: :  | | |:
Db     101 MDEMEFKPENSTRFPSKHLVNDISIFFPNGECN 133


RESULT 23
US-08-477-346-53
; Sequence 53, Application US/08477346
; Patent No. 6262023
;  GENERAL INFORMATION:
;    APPLICANT:  Mochly-Rosen, Daria
;    APPLICANT:  Ron, Dorit
;    TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  265
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,346
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/487,072
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  2550-0025.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  MSL1, Fig. 36
US-08-477-346-53

  Query Match             8.2%;  Score 69.5;  DB 4;  Length 422;
  Best Local Similarity   28.0%;  Pred. No. 20;
  Matches   26;  Conservative   18;  Mismatches   24;  Indels   25;  Gaps    6;

Qy      59 FP---TTSETPRVVYNSRT--DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYN 113
           ||   |||:  |:: :| |   ||    :|:: :|: |   ||  ||         ||::
Db      52 FPDLDTTSDEHRILLSSFTSSQKPEDETIYISKISTLGH--IKWSSL---------NNFD 100

Qy     114 SDDFQF-------VWNIYANNDVVV--PTGGCD 137
           |: :|         : :  ||: :  | | |:
Db     101 MDEMEFKPENSTRFPSKHLVNDISIFFPNGECN 133


RESULT 24
US-08-477-346-55
; Sequence 55, Application US/08477346
; Patent No. 6262023
;  GENERAL INFORMATION:
;    APPLICANT:  Mochly-Rosen, Daria
;    APPLICANT:  Ron, Dorit
;    TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  265
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,346
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/487,072
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  2550-0025.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  ORF RB1, Fig. 38
US-08-477-346-55

  Query Match             8.2%;  Score 69.5;  DB 4;  Length 422;
  Best Local Similarity   28.0%;  Pred. No. 20;
  Matches   26;  Conservative   18;  Mismatches   24;  Indels   25;  Gaps    6;

Qy      59 FP---TTSETPRVVYNSRT--DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYN 113
           ||   |||:  |:: :| |   ||    :|:: :|: |   ||  ||         ||::
Db      52 FPDLDTTSDEHRILLSSFTSSQKPEDETIYISKISTLGH--IKWSSL---------NNFD 100

Qy     114 SDDFQF-------VWNIYANNDVVV--PTGGCD 137
           |: :|         : :  ||: :  | | |:
Db     101 MDEMEFKPENSTRFPSKHLVNDISIFFPNGECN 133


RESULT 25
US-08-473-089-53
; Sequence 53, Application US/08473089
; Patent No. 6342368
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Mochly-Rosen, Daria
;    APPLICANT:  Ron, Dorit
;    TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  265
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/473,089
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  2550-0025.22
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  MSL1, Fig. 36
US-08-473-089-53

  Query Match             8.2%;  Score 69.5;  DB 4;  Length 422;
  Best Local Similarity   28.0%;  Pred. No. 20;
  Matches   26;  Conservative   18;  Mismatches   24;  Indels   25;  Gaps    6;

Qy      59 FP---TTSETPRVVYNSRT--DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYN 113
           ||   |||:  |:: :| |   ||    :|:: :|: |   || ||         ||::
Db      52 FPDLDTTSDEHRILLSSFTSSQKPEDETIYISKISTLGH--IKWSSL---------NNFD 100

Qy     114 SDDFQF-------VWNIYANNDVVV--PTGGCD 137
           |: :|         : :  ||: :  | | |:
Db     101 MDEMEFKPENSTRFPSKHLVNDISIFFPNGECN 133


RESULT 26
US-08-473-089-55
; Sequence 55, Application US/08473089
; Patent No. 6342368
;  GENERAL INFORMATION:
;    APPLICANT:  Mochly-Rosen, Daria
;    APPLICANT:  Ron, Dorit
;    TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  265
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/473,089
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  2550-0025.22
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  ORF RB1, Fig. 38
US-08-473-089-55

  Query Match             8.2%;  Score 69.5;  DB 4;  Length 422;
  Best Local Similarity   28.0%;  Pred. No. 20;
  Matches   26;  Conservative   18;  Mismatches   24;  Indels   25;  Gaps    6;

Qy      59 FP---TTSETPRVVYNSRT--DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYN 113
           ||   |||:  |:: :| |   ||    :|:: :|: |   || ||         ||::
Db      52 FPDLDTTSDEHRILLSSFTSSQKPEDETIYISKISTLGH--IKWSSL---------NNFD 100

Qy     114 SDDFQF-------VWNIYANNDVVV--PTGGCD 137
           |: :|         : :  ||: :  | | |:
Db     101 MDEMEFKPENSTRFPSKHLVNDISIFFPNGECN 133


RESULT 27
US-08-487-072A-53
; Sequence 53, Application US/08487072A
; Patent No. 6423684
;  GENERAL INFORMATION:
;    APPLICANT:  Mochly-Rosen, Daria
;    APPLICANT:  Ron, Dorit
;    TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  265
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,072A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  2550-0025.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;  INFORMATION FOR SEQ ID NO:  53:
;    SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  422 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  protein
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  MSL1, Fig. 36
US-08-487-072A-53

  Query Match             8.2%;  Score 69.5;  DB 4;  Length 422;
  Best Local Similarity   28.0%;  Pred. No. 20;
  Matches   26;  Conservative   18;  Mismatches   24;  Indels   25;  Gaps    6;

Qy      59 FP---TTSETPRVVYNSRT--DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYN 113
           ||    |||:  |:: :| |   ||    :|:: :|: |   ||  ||         ||::
Db      52 FPDLDTTSDEHRILLSSFTSSQKPEDETIYISKISTLGH--IKWSSL---------NNFD 100

Qy     114 SDDFQF-------VWNIYANNDVVV--PTGGCD 137
           |: :|         : :  ||: :  | | |:
Db     101 MDEMEFKPENSTRFPSKHLVNDISIFFPNGECN 133


RESULT 28
US-08-487-072A-55
; Sequence 55, Application US/08487072A
; Patent No. 6423684
;  GENERAL INFORMATION:
;    APPLICANT:  Mochly-Rosen, Daria
;    APPLICANT:  Ron, Dorit
;    TITLE OF INVENTION:  WD-40 - Derived Peptides and Uses
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  265
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Morrison & Foerster
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,072A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  2550-0025.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;  INFORMATION FOR SEQ ID NO:  55:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  ORF RB1, Fig. 38
US-08-487-072A-55

  Query Match             8.2%;  Score 69.5;  DB 4;  Length 422;
  Best Local Similarity   28.0%;  Pred. No. 20;
  Matches   26;  Conservative   18;  Mismatches   24;  Indels   25;  Gaps    6;

Qy      59 FP---TTSETPRVVYNSRT--DKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYN 113
           ||    |||:  |:: :| |   ||    :|:: :|: |   ||  ||         ||::
Db      52 FPDLDTTSDEHRILLSSFTSSQKPEDETIYISKISTLGH--IKWSSL---------NNFD 100

Qy     114 SDDFQF-------VWNIYANNDVVV--PTGGCD 137
           |: :|         : :  ||: :  | | |:
Db     101 MDEMEFKPENSTRFPSKHLVNDISIFFPNGECN 133


RESULT 29
US-09-331-581-23
; Sequence 23, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
;  APPLICANT: TOHDA, Hideki
;  APPLICANT: HAMA, Yuko
;  APPLICANT: KUMAGAI, Hiromicho
;  TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
;  TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
;  TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
;  FILE REFERENCE: 0059-1142-0PCT
;  CURRENT APPLICATION NUMBER: US/09/331,581
;  CURRENT FILING DATE: 1999-06-30
;  EARLIER APPLICATION NUMBER: PCT/JP98/04929
;  EARLIER FILING DATE: 1998-10-30
;  EARLIER APPLICATION NUMBER: JP 9-314608
;  EARLIER FILING DATE: 1997-10-31
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
;   LENGTH: 332
;   TYPE: PRT
;   ORGANISM: Schwanniomyces occidentalis
US-09-331-581-23

  Query Match             8.1%;  Score 69;  DB 4;  Length 332;
  Best Local Similarity   28.8%;  Pred. No. 16;
  Matches   38;  Conservative   12;  Mismatches   48;  Indels   34;  Gaps    9;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQR------GSA--YGGVL-SNFSGTVK 51
           ||::|  |   |   ::| |  :         : |      |||     || ||||
Db     158 PVIDVSSNQFRD--PKVFWHERFKSMDHGCSEIARVKIQIFGSANLKNWVLNSNFSS--G 213

Qy      52 YSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNN 111
           | |: |      | |  : ||   | | : | : | |  || :               |
Db     214 YYGNQYGMSRLIEVP--IENSDKSK-WVMFLAINPGSPLGGSI--------------NQ 255

Qy     112 Y---NSDDFQFV 120
           |   : | ||||
Db     256 YFVGDFDGFQFV 267


RESULT 30
US-07-723-002C-4
; Sequence 4, Application US/07723002C
; Patent No. 5447862
;  GENERAL INFORMATION:
;    APPLICANT:  Heim, Jutta
;    APPLICANT:  Meyhack, Bernd
;    APPLICANT:  Gysler, Christof
;    APPLICANT:  Visser, Jacob
;    APPLICANT:  Kester, Hermanus Cornelis Maria
;    TITLE OF INVENTION:  No. 5447862el Expression System
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/723,002C
;       FILING DATE:  28-JUN-1991
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 8702475
;       FILING DATE:  04-FEB-1987
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 150,880
;       FILING DATE:  29-JAN-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 8818046.8
;       FILING DATE:  28-JUL-1988
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  GB 8914666.6
;       FILING DATE:  26-JUN-1989
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 384,898
;       FILING DATE:  24-JUL-1989
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Foley, Shawn P.
;       REGISTRATION NUMBER:  33,071
;       REFERENCE/DOCKET NUMBER:  4-16317/+/CIP
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (919)541-8615
;       TELEFAX:  (919)541-8689
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  379 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-07-723-002C-4

  Query Match             8.1%;  Score 69;  DB 1;  Length 379;
  Best Local Similarity   22.7%;  Pred. No. 19;
  Matches   34;  Conservative   21;  Mismatches   69;  Indels   26;  Gaps    6;

Qy      21 NDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPV 80
           |:| : ::||     |  | |:  :     :     :| : |:  :|:|  |:       |
Db     213 NNYIDGVSDYSATCDGYHYWGIYLDGDADLVTMKGNYIYHTSGRSPKVQDNTLLH---CV 269

Qy      81 ALYLTPVS------SAGGLVIKAGSLI----AVLILRQTNNYNSDDFQF-------VWNI 123
             |   :|        || |:  |::       ||       |    |         | :
Db     270 NNYFYDISGHAFEIGEGGYVLAEGNVFQNVDTVL-----ETYEGAAFTVPSTTAGEVCST 324

Qy     124 YANNDVVVPTGGCD-VSARDVTVTLPDYRG 152
           |   | |:   ||    : | |  | |: |
Db     325 YLGRDCVINGFGCSGTFSEDSTSFLSDFEG 354


RESULT 31
US-08-712-072C-2
; Sequence 2, Application US/08712072C
; Patent No. 5925541
;  GENERAL INFORMATION:
;    APPLICANT:  Jack Goldstein, Alex Zhu and Lin Leng
;    TITLE OF INVENTION:  ENDO-BETA-GALACTOSIDASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amster, Rothstein & Ebenstein
;      STREET:  90 Park Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  US
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/712,072C
;      FILING DATE:  11-SEP-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bogosian, Elizabeth A.
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  63475/97
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  422 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
US-08-712-072C-2

  Query Match             8.1%;  Score 69;  DB 2;  Length 422;
  Best Local Similarity   22.6%;  Pred. No. 22;
  Matches   48;  Conservative   23;  Mismatches   67;  Indels   74;  Gaps   11;

Qy      12 DLSTQIFCHND----------------------YPETITDYVTLQR--GSAYGGVLSN- 45
           :| | |:  ||                      :|  :  |: | :  |: : | ::|
Db     214 NLYTMIWSPNDIRFYVNNSLQYTYARVSGGGTQQWPFDVPFYLILNQAGGAGWPGAITNA 273

Qy      46 ---FSGTVKY--------------SGSSYPFPTTSETPRVVYNS---------RTDKPW 78
              ||  | |               ||  ||: |      ||              |
Db     274 DLPFSMQVDYVRVYKLPLFSNGDFESGVIYPWTTWGGGSSVVSTDARTGTKCIRETGGET 333

Qy      79 PVALYLTPVS-----SAGGL--VIKAGSLIAVLILRQTNNYNSD--DFQFVWNIYANNDV 129
           :  ||| ::       ||   |  ||  ::: :     ||     |       |:|| |
Db     334 SIEQYLTGLTPNTTYRFGGYAKVSAAGQSVSIGV----KNYGGTAVDATIGTTSYSNNSV 389

Qy     130 VVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||     | : | |:  |:   |:  |||
Db     390 TFTTG-----ANNTTATVYFYK---PLSGTVY 413


RESULT 32
US-09-620-412C-309
; Sequence 309, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
;  APPLICANT: Steven P. Fling
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
;  TITLE OF INVENTION:  DIAGNOSIS OF CHLAMYDIAL INFECTION
;  FILE REFERENCE: 210121.469C7
;  CURRENT APPLICATION NUMBER: US/09/620,412C
;  CURRENT FILING DATE:  2000-07-20
;  NUMBER OF SEQ ID NOS: 363
;  SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 309
;   LENGTH: 619
;   TYPE: PRT
;   ORGANISM: Chlamydia trachomatis
US-09-620-412C-309

  Query Match             8.1%;  Score 69;  DB 4;  Length 619;
  Best Local Similarity   21.3%;  Pred. No. 39;
  Matches   27;  Conservative   21;  Mismatches   61;  Indels   18;  Gaps    5;

Qy      31 VTLQRGSA-YGGVLSNFSGTVKYSGSSYPFP----TTSETPRVVYNSR---TDKPWPVAL 82
```

```
          :| :  || :|| : | :| |:: |:: |      ||      :| |      :  |:
Db    402 ITFEGNSAVHGGAIYNKNGLVEFLGNAGPLAFKENTTIANGGAIYTSNFKANQQTSPILF 461

Qy     83 YLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD 142
             :  ||         |:         | |  :|  |        : :  | ::| :
Db    462 SQNHANKKGG---------AIYAQYVNLEQNQDTIRFEKNTAKEGGGAITSSQCSITAHN 512

Qy    143 VTVTLPD 149
           |:|  |
Db    513 -TITFSD 518


RESULT 33
US-09-071-035-198
; Sequence 198, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  198:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  129 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-198

  Query Match             8.1%;  Score 68.5;  DB 4;  Length 129;
  Best Local Similarity   24.7%;  Pred. No. 4.6;
  Matches   20;  Conservative   10;  Mismatches   30;  Indels   21;  Gaps    3;

Qy     18 FCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKP 77
          |||:|              :|:|      || | : | :| :||   |     :| | :|
Db     16 FCHSD-------------TAFGEAAYENSGVVSFYG-TYEYPTEESTTATSNSSTTTEP 60

Qy     78 WPVALYLTPVSSAGGLVIKAG 98
                  |       |: :|
Db     61 ------TKPADGGASSVLSSG 75


RESULT 34
US-09-216-295-20
; Sequence 20, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
;  APPLICANT:  Mitchinson, Colin
;  APPLICANT:  Wendt, Dan J.
;  TITLE OF INVENTION:  No. 6268328el Variant EGIII-Like Cellulase Compositions
;  FILE REFERENCE:  GC555
;  CURRENT APPLICATION NUMBER: US/09/216,295
;  CURRENT FILING DATE:  1998-12-18
;  NUMBER OF SEQ ID NOS:  41
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 20
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Emericella desertoru
US-09-216-295-20

  Query Match             8.1%;  Score 68.5;  DB 4;  Length 246;
  Best Local Similarity   21.9%;  Pred. No. 12;
  Matches   35;  Conservative   20;  Mismatches   60;  Indels   45;  Gaps    6;

Qy     38 AYGGVLSNFSGTVKYSGSSYP------FPTTSETPRVVYNSRT------DKPWPVALYLT 85
          :|      |: |   | || |      : ||     | |:  |      |  : : ::|
Db     88 SYANAAYQFTSTKLNSLSSIPTSWKWQYSTTDIVANVAYDLFTSSSAGGDSEYEIMIWLA 147

Qy     86 PVSSAGGLVIKAGSLIAVLIL----------------------RQTNNYNSDDFQFVWN 122
           :  ||  :   || || : |                      | ::::|   |: |
Db    148 ALGGAGP-ISSTGSSIATVTLGGVTWSLYSGPNGSMQVYSFVASSTTESFSADLMDFI-N 205

Qy    123 IYANNDVV--------VPTGGCDVSARDVTVTLPDYRGSV 154
             | |  :        |  |    :  | |:|:  |  ||
Db    206 YLAENQGLSSSQYLTHVQAGTEPFTGTDATLTVSSYSVSV 245


RESULT 35
US-09-032-523-2
; Sequence 2, Application US/09032523
; Patent No. 6232454
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl
;    APPLICANT:  Baugh, Mariah
;    TITLE OF INVENTION:  HUMAN PROTEINASE MOLECULES
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/032,523
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0479 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  415 amino acids
```

```
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  RATRNOT02
;      CLONE:  947429
US-09-032-523-2

  Query Match             8.1%;  Score 68.5;  DB 4;  Length 415;
  Best Local Similarity   19.7%;  Pred. No. 25;
  Matches   35;  Conservative   20;  Mismatches   56;  Indels   67;  Gaps    6;

Qy     12 DLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYN 71
          |:    :|  ||                 || :::     || | | | |  ||   ::
Db    201 DVERDNYCRYDYVAVFN----------GGEVNDARRIGKYCGDSPPAPIVSERNELLIQ 249

Qy     72 ------------------------SRTDKP----WPVALYLTP-VSSAGGLVIKAGSLIA 102
                                    : |::|    :||   | | |:       : |:|
Db    250 FLSDLSLTADGFIGHYIFRPKKLPTTTEQPVTTTFPVTTGLKPTVALCQQKCRRTGTL-- 307

Qy    103 VLILRQTNNYNSDDFQF------------------VWNIYANNDVVVPTGGCDVSAR 141
              || | ||                        : |||   :: :   | ::|||
Db    308 ------EGNYCSSDFVLAGTVITTITRDGSLHATVSIINIYKEGNLAIQQAGKNMSAR 359


RESULT 36
US-08-947-965-73
; Sequence 73, Application US/08947965A
; Patent No. 6004790
; GENERAL INFORMATION:
;  APPLICANT: Dijkhuizen, Lubbert
;  APPLICANT: Dijkstra, Bauke
;  APPLICANT: Andersen, Carsten
;  APPLICANT: Osten, Claus von der
;  TITLE OF INVENTION: Cyclomaltodextrin Glucanotransferase
;  TITLE OF INVENTION: Variants
;  FILE REFERENCE: 4285.204-US
;  CURRENT APPLICATION NUMBER: US/08/947,965A
;  CURRENT FILING DATE: 1997-10-09
;  EARLIER APPLICATION NUMBER: 0477/95
;  EARLIER FILING DATE: 1995-04-21
;  EARLIER APPLICATION NUMBER: 1173/95
;  EARLIER FILING DATE: 1995-10-17
;  EARLIER APPLICATION NUMBER: 1281/95
;  EARLIER FILING DATE: 1995-11-16
;  EARLIER APPLICATION NUMBER: PCT/DK96/00179
;  EARLIER FILING DATE: 1996-04-22
;  NUMBER OF SEQ ID NOS: 78
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 73
;   LENGTH: 686
;   TYPE: PRT
;   ORGANISM: Bacillus sp.
US-08-947-965-73

  Query Match             8.1%;  Score 68.5;  DB 3;  Length 686;
  Best Local Similarity   24.5%;  Pred. No. 51;
  Matches   26;  Conservative   20;  Mismatches   37;  Indels   23;  Gaps    4;

Qy      2 VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAY----GGVLSNFSGTVKYSGSSY 57
            || : :|:            : | :||  ||  | ::|    ||:|:  : ||   |::
Db    431 VVAINRNM-----------NTPASITGLVTSLRRASYNDVLGGILNGNTLTVGAGGAAS 478

Qy     58 PFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAV 103
           | | :     |:   ||   |:      : : | :: | |  | :
Db    479 NF-TLAPGGTAVWQYTTDATTPI------IGNVGPMMAKPGVTITI 517


RESULT 37
5194600-2
;Patent No. 5194600
;    APPLICANT: BUSSEY, HOWARD;BOONE, CHARLES;SOMMER, STEVE S.;
;HILL, KATHRYN;MEADEN, PHILIP
;    TITLE OF INVENTION: GENES WHICH PARTICIPATE IN B-GLUCAN
;ASSEMBLY AND USE THEREOF
;    NUMBER OF SEQUENCES: 4
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,316
;      FILING DATE: 05-MAR-1990
;SEQ ID NO:2:
;      LENGTH: 291
5194600-2

  Query Match             8.0%;  Score 68;  DB 6;  Length 291;
  Best Local Similarity   26.2%;  Pred. No. 17;
  Matches   39;  Conservative   23;  Mismatches   67;  Indels   20;  Gaps    7;

Qy     12 DLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYN 71
          || | :         |:|   |    ||      |: | :|  ||||    ||: |   |
Db    109 DLGTTV--------TLTQTFTHSSTSATSSASSSVSSSVSSSGSSSSVKTTTSTGSAVAE 160

Qy     72 SRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQ--TNNY---NSDDFQFVWNIYAN 126
          : | :| |   :  |  ||    : : |: : : : :  |: |    :    :|  :
Db    161 TGT-RPDPSTDFTEPPVSA----VTSLSIDSYITITEGTTSTYTTTRAPTSMWVTVVRQG 215

Qy    127 NDVVVPTGGCD-VSARDVTVTLPDYRGSV 154
          | : | |      |:: |||  |   ||:
Db    216 NTITVQTTFVQRFSSQYVTVASPSV-GSI 243


RESULT 38
US-09-071-035-244
; Sequence 244, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  244:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  624 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-244

  Query Match             8.0%;  Score 68;  DB 4;  Length 624;
  Best Local Similarity   21.1%;  Pred. No. 51;
  Matches   42;  Conservative   24;  Mismatches   61;  Indels   72;  Gaps    8;
```

```
Qy      11 VDLSTQIFCHNDYPET-----------------ITDYVTLQ----RGS----------- 37
           |||:| :   :|| ||                  ||:  |:     :|:
Db     321 VDLATGVSFFDDYDETXVTPIKDLLRVKDSKGXDITNQFTISWDDAKGTVTXSAKDPQAF 380

Qy      38 --AYGG----------VLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWP---VAL 82
             | ||          | :: || |  |      |    :|  ||  :   |  |   | :
Db     381 ILAXGGQELRVTLPTKVKADVSGDVYNSAEQNTFGQRIKTNTVV--NHIPKVXPKKDVVI 438

Qy      83 YLTPVSSAGGLVIKAG----------------SLIAVLILRQTNNYNSDDFQFVWNIYA 125
           :    |  |  || |                 ::    :    :   | |   |:::|
Db     439 KVGDKQSQNGATIKLGEXFFYEFTSSDIPAEYAGVVEEWSISDKLDVKHDKFSGQWSVFA 498

Qy     126 NNDVVVPTG-----GCDVS 139
           |:: |:  |     | |:|
Db     499 NSNFVLADGTKVNKGDDIS 517


RESULT 39
US-08-816-105A-1
; Sequence 1, Application US/08816105A
; Patent No. 5989882
;  GENERAL INFORMATION:
;    APPLICANT:  Crueger, Anneliese; Dellweg, Hans-Georg; Lenz,
;    APPLICANT:  Jurgen; Schroder, Werner; Pape, Hermann;
;    APPLICANT:  Goeke, Klaus; Schaper, Beate; Hemker, Michael;
;    APPLICANT:  Piepersberg, Wolfgang; Distler, Jurgen;
;    APPLICANT:  Stratmann, Ansgar
;    TITLE OF INVENTION:  PROCESSES FOR PREPARING ACARVIOSYL
;    TITLE OF INVENTION:  TRANSFERASE AND FOR USING IT IN THE
;    TITLE OF INVENTION:  CONVERSION OF ACARBOSE HOMOLOGUES
;    TITLE OF INVENTION:  INTO ACARBOSE, FOR THE PREPARATION
;    TITLE OF INVENTION:  OF ACARBOSE HOMOLOGUES
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;      COMPUTER:  Apple Macintosh 6500
;      OPERATING SYSTEM:  System 7.5
;      SOFTWARE:  WordPerfect 3.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/816,105A
;      FILING DATE:  14-MAR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19625269.5
;      FILING DATE:  25-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19611252.4
;      FILING DATE:  22-MAR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 9814-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  725 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-816-105A-1

  Query Match             8.0%;  Score 68;  DB 2;  Length 725;
  Best Local Similarity   24.5%;  Pred. No. 63;
  Matches   40;  Conservative   13;  Mismatches   36;  Indels   74;  Gaps    9;

Qy      24 PETITDY------VTLQRGSAYGGV---LSNFSGTVKYSGSSYPFPTTSETPRVVYNSRT 74
           | |:  :       |:  |   | |   : | |||     : ||  |                ||
Db     575 PATVQQWTADRITATVPVGVPTGAVQVTVGNGSGT----SNGYPITT----------RT 619

Qy      75 DKPWPV-------------ALYLT-----------PVSSAGGLVIKAGSLIAVLILRQ 108
           || ||               :||||           |  :|| |:        ||:
Db     620 GKPVPVQFTVQNPPATAPGESLYLTGDVAELGHWSTSPDQTAGQLLRVPNESRGVLV--- 676

Qy     109 TNNYNSD-------DFQFV---------WNIYANNDVVVPTGG 135
               :|       :|:||         |   ||:   || ||
Db     677 -----ADLPAGAPVEFKFVKVAADGTVTWEGGANHRYTVPAGG 714


RESULT 40
US-09-071-035-236
; Sequence 236, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  236:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1223 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-236

  Query Match             8.0%;  Score 68;  DB 4;  Length 1223;
  Best Local Similarity   21.1%;  Pred. No. 1.3e+02;
  Matches   42;  Conservative   24;  Mismatches   61;  Indels   72;  Gaps    8;

Qy      11 VDLSTQIFCHNDYPET-----------------ITDYVTLQ----RGS----------- 37
           |||:| :   :|| ||                  ||:  |:     :|:
Db     920 VDLATGVSFFDDYDETXVTPIKDLLRVKDSKGXDITNQFTISWDDAKGTVTXSAKDPQAF 979

Qy      38 --AYGG----------VLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWP---VAL 82
             | ||          | :: || |  |      |    :|  ||  :   |  |   | :
Db     980 ILAXGGQELRVTLPTKVKADVSGDVYNSAEQNTFGQRIKTNTVV--NHIPKVXPKKDVVI 1037

Qy      83 YLTPVSSAGGLVIKAG----------------SLIAVLILRQTNNYNSDDFQFVWNIYA 125
           :    |  |  || |                 ::    :    :   | |   |:::|
Db    1038 KVGDKQSQNGATIKLGEXFFYEFTSSDIPAEYAGVVEEWSISDKLDVKHDKFSGQWSVFA 1097
```

```
Qy    126 NNDVVVPTG-----GCDVS 139
          |:: |:  |     | |:|
Db   1098 NSNFVLADGTKVNKGDDIS 1116


RESULT 41
US-09-071-035-234
; Sequence 234, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  234:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1301 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-234

  Query Match             8.0%;  Score 68;  DB 4;  Length 1301;
  Best Local Similarity   21.1%;  Pred. No. 1.5e+02;
  Matches   42;  Conservative   24;  Mismatches   61;  Indels   72;  Gaps    8;

Qy     11 VDLSTQIFCHNDYPET-----------------ITDYVTLQ----RGS----------- 37
          |||:| :   :|| ||                  ||:  |:     :|:
Db    959 VDLATGVSFFDDYDETXVTPIKDLLRVKDSKGXDITNQFTISWDDAKGTVTXSAKDPQAF 1018

Qy     38 --AYGG----------VLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWP---VAL 82
            | ||          | :: || |  |     |   :|  ||  :   |  |   | :
Db   1019 ILAXGGQELRVTLPTKVKADVSGDVYNSAEQNTFGQRIKTNTVV--NHIPKVXPKKDVVI 1076

Qy     83 YLTPVSSAGGLVIKAG----------------SLIAVLILRQTNNYNSDDFQFVWNIYA 125
          :    |  |  || |                 ::    :    :   | |   |:::|
Db   1077 KVGDKQSQNGATIKLGEXFFYEFTSSDIPAEYAGVVEEWSISDKLDVKHDKFSGQWSVFA 1136

Qy    126 NNDVVVPTG-----GCDVS 139
          |:: |:  |     | |:|
Db   1137 NSNFVLADGTKVNKGDDIS 1155


RESULT 42
US-09-071-035-238
; Sequence 238, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  238:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1301 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-238

  Query Match             8.0%;  Score 68;  DB 4;  Length 1301;
  Best Local Similarity   21.1%;  Pred. No. 1.5e+02;
  Matches   42;  Conservative   24;  Mismatches   61;  Indels   72;  Gaps    8;

Qy     11 VDLSTQIFCHNDYPET-----------------ITDYVTLQ----RGS----------- 37
          |||:| :   :|| ||                  ||:  |:     :|:
Db    959 VDLATGVSFFDDYDETXVTPIKDLLRVKDSKGXDITNQFTISWDDAKGTVTXSAKDPQAF 1018

Qy     38 --AYGG----------VLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWP---VAL 82
            | ||          | :: || |  |     |   :|  ||  :   |  |   | :
Db   1019 ILAXGGQELRVTLPTKVKADVSGDVYNSAEQNTFGQRIKTNTVV--NHIPKVXPKKDVVI 1076

Qy     83 YLTPVSSAGGLVIKAG----------------SLIAVLILRQTNNYNSDDFQFVWNIYA 125
          :    |  |  || |                 ::    :    :   | |   |:::|
Db   1077 KVGDKQSQNGATIKLGEXFFYEFTSSDIPAEYAGVVEEWSISDKLDVKHDKFSGQWSVFA 1136

Qy    126 NNDVVVPTG-----GCDVS 139
          |:: |:  |     | |:|
Db   1137 NSNFVLADGTKVNKGDDIS 1155


RESULT 43
US-09-071-035-242
; Sequence 242, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
```

```
;       COUNTRY:  USA
;       ZIP:  20850
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;       COMPUTER:  HP Vectra 486/33
;       OPERATING SYSTEM:  MSDOS version 6.2
;       SOFTWARE:  ASCII Text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/071,035
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  A. Anders Brookes
;       REGISTRATION NUMBER:  36,373
;       REFERENCE/DOCKET NUMBER:  PB369P2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (301) 309-8504
;       TELEFAX:  (301) 309-8512
;   INFORMATION FOR SEQ ID NO:  242:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1301 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-09-071-035-242

  Query Match             8.0%;  Score 68;  DB 4;  Length 1301;
  Best Local Similarity   21.1%;  Pred. No. 1.5e+02;
  Matches   42;  Conservative   24;  Mismatches   61;  Indels   72;  Gaps    8;

Qy      11 VDLSTQIFCHNDYPET------------------ITDYVTLQ----RGS----------- 37
           |||:| :   :|| ||                  ||:  |:     :|:
Db     959 VDLATGVSFFDDYDETXVTPIKDLLRVKDSKGXDITNQFTISWDDAKGTVTXSAKDPQAF 1018

Qy      38 --AYGG----------VLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWP---VAL 82
             | ||          | :: || |  |      |    :|  ||  :   |  |   | :
Db    1019 ILAXGGQELRVTLPTKVKADVSGDVYNSAEQNTFGQRIKTNTVV--NHIPKVXPKKDVVI 1076

Qy      83 YLTPVSSAGGLVIKAG----------------SLIAVLILRQTNNYNSDDFQFVWNIYA 125
            :   |  |  || |                ::   :    :  | |   |:::|
Db    1077 KVGDKQSQNGATIKLGEXFFYEFTSSDIPAEYAGVVEEWSISDKLDVKHDKFSGQWSVFA 1136

Qy     126 NNDVVVPTG-----GCDVS 139
           |:: |:  |     | |:|
Db    1137 NSNFVLADGTKVNKGDDIS 1155


RESULT 44
US-08-612-734B-2
; Sequence 2, Application US/08612734B
; Patent No. 5914246
;  GENERAL INFORMATION:
;    APPLICANT:  Peery, Robert B.
;    APPLICANT:  Skatrud, Paul L.
;    APPLICANT:  Tobin, Matthew B.
;    TITLE OF INVENTION:  Multiple Drug Resistance Gene of
;    TITLE OF INVENTION:  Aspergillus Fumigatus
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center, DC1501
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  U.S.A.
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/612,734B
;      FILING DATE:  08-MAR-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Craig, Anne I.
;      REGISTRATION NUMBER:  32,976
;      REFERENCE/DOCKET NUMBER:  X-9681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-354-9570
;      TELEFAX:  617-354-4043
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1349 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-612-734B-2

  Query Match             8.0%;  Score 68;  DB 2;  Length 1349;
  Best Local Similarity   25.3%;  Pred. No. 1.6e+02;
  Matches   42;  Conservative   17;  Mismatches   61;  Indels   46;  Gaps    7;

Qy       3 VNVGQNLVVDLST-------------QIFCHNDYPETITDYVTLQRGSAY------GGV 42
           ||||| | | :|               | |  |        | |: | |        | |
Db     371 VNVGQVLTVLMSILIGSFSLGNVAPNGQAFT-NGVAAAAKIYSTIDRRSPLDPYSDEGKV 429

Qy      43 LSNFSGTVKYSGSSYPFPTTSE-------------TPRVVYNSRTDKPWPVAL---YL 84
           | :| | :::    : :|:  |               |  :|  | : |   | |   :
Db     430 LDHFEGNIEFRNVKHIYPSRPEVTVMEDVSLSMPAGKTTALVGPSGSGKSTVVGLVERFY 489

Qy      85 TPVSSAGGLVIKAGSLIAVLILR----QTNNYNSDDFQFVWNIYAN 126
            ||   || |:  |  |  | ||    | :  : :   |   |: |
Db     490 LPV---GGQVLLDGHDIQTLNLRWLRQQISLVSQEPVLFSTTIFRN 532


RESULT 45
US-09-320-878-2
; Sequence 2, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-2

  Query Match             8.0%;  Score 68;  DB 3;  Length 3739;
  Best Local Similarity   23.4%;  Pred. No. 6.8e+02;
```

Matches 41; Conservative 25; Mismatches 61; Indels 48; Gaps 8;

```
Qy      24 PETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPF----PTTSETPRV-VYNSRTDKPW 78
           |||:|   ||:| :     |:        :| :  :    |||:  | :  |  :|:: |
Db    2348 PETVTGLGTLRRDNGGQHRLTTSLAEAWANGLTVDWASLLPTTTTHPDLPTYAFQTERYW 2407

Qy      79 PVALYLTPVSSAGGLVIKA-------------------GSLIAVLILRQTNNYNSD--- 115
           |      |  || | :  |                   | |:   :  :|: : :|
Db    2408 P-----QPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAV 2462

Qy     116 -------DFQFVWNIYANNDVVVPTGGCDVSAR---DVTVTLPDYRGSVPIPLTV 160
                  ||   :   | |    |||:       |  : ||  ||:| : |:|
Db    2463 AGTVLLPGTAFVELAFRAGDQV----GCDLVEELTLDAPLVLP-RRGAVRVQLSV 2512
```

Search completed: November 28, 2002, 18:54:45
Job time : 30 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 17:21:36 ; Search time 19 Seconds
(without alignments)
134.937 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	848	100.0	279	10	US-09-900-575-29	Sequence 29, Appl
2	845	99.6	279	10	US-09-900-575-36	Sequence 36, Appl
3	838	98.8	279	10	US-09-900-575-42	Sequence 42, Appl
4	837	98.7	279	10	US-09-900-575-55	Sequence 55, Appl
5	834	98.3	279	10	US-09-900-575-28	Sequence 28, Appl
6	834	98.3	279	10	US-09-900-575-32	Sequence 32, Appl
7	834	98.3	279	10	US-09-900-575-34	Sequence 34, Appl
8	834	98.3	279	10	US-09-900-575-44	Sequence 44, Appl
9	834	98.3	300	10	US-09-912-020-367	Sequence 367, App
10	833	98.2	279	10	US-09-900-575-37	Sequence 37, Appl
11	832	98.1	279	10	US-09-900-575-27	Sequence 27, Appl
12	831	98.0	279	10	US-09-900-575-23	Sequence 23, Appl
13	831	98.0	279	10	US-09-900-575-43	Sequence 43, Appl
14	829	97.8	279	10	US-09-900-575-38	Sequence 38, Appl
15	829	97.8	279	10	US-09-900-575-40	Sequence 40, Appl
16	827	97.5	279	10	US-09-900-575-39	Sequence 39, Appl
17	827	97.5	280	10	US-09-900-575-30	Sequence 30, Appl
18	826	97.4	279	10	US-09-900-575-25	Sequence 25, Appl
19	826	97.4	279	10	US-09-900-575-26	Sequence 26, Appl
20	826	97.4	279	10	US-09-900-575-45	Sequence 45, Appl
21	825	97.3	279	10	US-09-900-575-31	Sequence 31, Appl
22	823	97.1	279	10	US-09-900-575-35	Sequence 35, Appl
23	819	96.6	279	10	US-09-900-575-24	Sequence 24, Appl
24	819	96.6	279	10	US-09-900-575-33	Sequence 33, Appl
25	819	96.6	279	10	US-09-900-575-41	Sequence 41, Appl
26	83	9.8	499	10	US-09-732-350-3	Sequence 3, Appli
27	75.5	8.9	358	10	US-09-815-242-12939	Sequence 12939, A
28	75.5	8.9	448	10	US-09-815-242-5806	Sequence 5806, Ap
29	75.5	8.9	451	10	US-09-815-242-13127	Sequence 13127, A
30	71	8.4	410	10	US-09-804-551B-18	Sequence 18, Appl
31	70.5	8.3	471	10	US-09-881-752A-18	Sequence 18, Appl
32	70.5	8.3	511	10	US-09-815-242-13697	Sequence 13697, A
33	69	8.1	184	9	US-09-741-233A-6	Sequence 6, Appli
34	69	8.1	185	10	US-09-970-638-5	Sequence 5, Appli
35	69	8.1	245	10	US-09-995-598-24	Sequence 24, Appl
36	69	8.1	619	10	US-09-841-132-309	Sequence 309, App
37	69	8.1	5701	10	US-09-864-761-37319	Sequence 37319, A
38	69	8.1	5795	10	US-09-815-242-12610	Sequence 12610, A
39	68.5	8.1	239	10	US-09-995-598-50	Sequence 50, Appl
40	68.5	8.1	244	10	US-09-995-598-12	Sequence 12, Appl
41	68.5	8.1	415	9	US-09-905-291A-104	Sequence 104, App
42	68.5	8.1	415	10	US-09-909-320-104	Sequence 104, App
43	68.5	8.1	415	10	US-09-909-088B-104	Sequence 104, App
44	68.5	8.1	922	10	US-09-886-468-19	Sequence 19, Appl
45	68	8.0	3739	9	US-09-860-846-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-09-900-575-29
; Sequence 29, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-29

Query Match 100.0%; Score 848; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.8e-83;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
```

RESULT 2
US-09-900-575-36

```
; Sequence 36, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-36

  Query Match             99.6%;  Score 845;  DB 10;  Length 279;
  Best Local Similarity   99.4%;  Pred. No. 3.7e-83;
  Matches  160;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVVIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 3
US-09-900-575-42
; Sequence 42, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 42
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-42

  Query Match             98.8%;  Score 838;  DB 10;  Length 279;
  Best Local Similarity   98.1%;  Pred. No. 2.1e-82;
  Matches  158;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: |||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 4
US-09-900-575-55
; Sequence 55, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 55
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Consensus sequence of FimH proteins for SEQ ID NO: 23 to 45
US-09-900-575-55

  Query Match             98.7%;  Score 837;  DB 10;  Length 279;
  Best Local Similarity   98.1%;  Pred. No. 2.6e-82;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||: |||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 5
US-09-900-575-28
; Sequence 28, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-28

  Query Match             98.3%;  Score 834;  DB 10;  Length 279;
  Best Local Similarity   99.4%;  Pred. No. 5.5e-82;
  Matches  158;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy     63 SETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWN 122
          ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db     88 SETPRVVYNSRTDKPWPVALYLTPVSSAGGVVIKAGSLIAVLILRQTNNYNSDDFQFVWN 147

Qy    123 IYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          |||||||||||||||||||||||||||||||||||||||
Db    148 IYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 6
US-09-900-575-32
; Sequence 32, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-32

  Query Match             98.3%;  Score 834;  DB 10;  Length 279;
  Best Local Similarity   98.1%;  Pred. No. 5.5e-82;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVVIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          |||||||||||||||||||||||||||||| |||||||||
Db    146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 186


RESULT 7
US-09-900-575-34
; Sequence 34, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-34

  Query Match             98.3%;  Score 834;  DB 10;  Length 279;
  Best Local Similarity   98.1%;  Pred. No. 5.5e-82;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||||||||||||||||||||||||||||: |||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          |||||||||||||||||||||||||||||| |||||||||
Db    146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 186


RESULT 8
US-09-900-575-44
; Sequence 44, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-44

  Query Match             98.3%;  Score 834;  DB 10;  Length 279;
  Best Local Similarity   98.1%;  Pred. No. 5.5e-82;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          ||||||||||||||||||||||||||||||: |||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          |||||||||||||||||||||||||||||| |||||||||
Db    146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 186


RESULT 9
US-09-912-020-367
; Sequence 367, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Trawick, John
;  APPLICANT:  Forsyth, R. Allyn
;  APPLICANT:  Froelich, Jamie M.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
```

```
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001DV1
;  CURRENT APPLICATION NUMBER: US/09/912,020
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 367
;   LENGTH: 300
;   TYPE: PRT
;   ORGANISM: E. Coli
US-09-912-020-367

  Query Match             98.3%;  Score 834;  DB 10;  Length 300;
  Best Local Similarity   98.1%;  Pred. No. 6.1e-82;
  Matches  158;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 166

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||||||||||||||||| ||||||||||
Db     167 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 207


RESULT 10
US-09-900-575-37
; Sequence 37, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 37
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-37

  Query Match             98.2%;  Score 833;  DB 10;  Length 279;
  Best Local Similarity   97.5%;  Pred. No. 7.1e-82;
  Matches  157;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVEYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 11
US-09-900-575-27
; Sequence 27, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-27

  Query Match             98.1%;  Score 832;  DB 10;  Length 279;
  Best Local Similarity   97.5%;  Pred. No. 9.1e-82;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||| ||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSAHDVTVTLPDYRGSVPIPLTVY 186


RESULT 12
US-09-900-575-23
; Sequence 23, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-23

  Query Match             98.0%;  Score 831;  DB 10;  Length 279;
  Best Local Similarity   97.5%;  Pred. No. 1.2e-81;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGIVKYSGSSYPFP 85
```

```
Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||||: |||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||| ||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDASARDVTVTLPDYRGSVPIPLTVY 186


RESULT 13
US-09-900-575-43
; Sequence 43, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 43
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-43

  Query Match             98.0%;  Score 831;  DB 10;  Length 279;
  Best Local Similarity   97.5%;  Pred. No. 1.2e-81;
  Matches  157;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: ||||||||||||||| |||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTKNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 14
US-09-900-575-38
; Sequence 38, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-38

  Query Match             97.8%;  Score 829;  DB 10;  Length 279;
  Best Local Similarity   96.9%;  Pred. No. 1.9e-81;
  Matches  156;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           . | |||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSHFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: ||||||:||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLMAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 15
US-09-900-575-40
; Sequence 40, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-40

  Query Match             97.8%;  Score 829;  DB 10;  Length 279;
  Best Local Similarity   96.9%;  Pred. No. 1.9e-81;
  Matches  156;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 16
US-09-900-575-39
; Sequence 39, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 39
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-39

  Query Match             97.5%;  Score 827;  DB 10;  Length 279;
  Best Local Similarity   96.9%;  Pred. No. 3.1e-81;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          |||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          ||||||||||||||||| ||||| ||||||||||||||||
Db    146 WNIYANNDVVVPTGGCDVSVRDVTVILPDYRGSVPIPLTVY 186


RESULT 17
US-09-900-575-30
; Sequence 30, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
;   LENGTH: 280
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-30

  Query Match             97.5%;  Score 827;  DB 10;  Length 280;
  Best Local Similarity   96.9%;  Pred. No. 3.1e-81;
  Matches  156;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          | |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     26 PAVNVGQNLVVDLSTQTFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          |||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          ||||||||||||||||| ||||||||||||||||||
Db    146 WNIYANNDVVVPTGGCDVSAHDVTVTLPDYRGSVPIPLTVY 186


RESULT 18
US-09-900-575-25
; Sequence 25, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-25

  Query Match             97.4%;  Score 826;  DB 10;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 4e-81;
  Matches  155;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          | |||||||||||||||||||||||||||||||||:|||||||:|||||||:|||||||
Db     26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          |||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          |||||||||||||||||||||||||||||||||||||||||||
Db    146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 19
US-09-900-575-26
; Sequence 26, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-26

  Query Match             97.4%;  Score 826;  DB 10;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 4e-81;
  Matches  155;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          | |||||||||||||||||||||||||||||||||:|||||||:|||||||:|||||||
Db     26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
          |||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy    121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
          |||||||||||||||||||||||||||||||||||||||||||
Db    146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 20
US-09-900-575-45
; Sequence 45, Application US/09900575
```

```
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 45
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-45

  Query Match             97.4%;  Score 826;  DB 10;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 4e-81;
  Matches  155;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 21
US-09-900-575-31
; Sequence 31, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-31

  Query Match             97.3%;  Score 825;  DB 10;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 5.1e-81;
  Matches  155;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||| |||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDASARDVTVTLPDYRGSVPIPLTVY 186


RESULT 22
US-09-900-575-35
; Sequence 35, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-35

  Query Match             97.1%;  Score 823;  DB 10;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 8.4e-81;
  Matches  155;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||| ||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGSVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 23
US-09-900-575-24
; Sequence 24, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-24

  Query Match             96.6%;  Score 819;  DB 10;  Length 279;
  Best Local Similarity   95.7%;  Pred. No. 2.2e-80;
  Matches  154;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | |||||||||||||||||||||||||||||||||||||||||:||| |||:|||||||
```

```
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGIVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||||: ||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||| |||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDASARDVTVTLPDYRGSVPIPLTVY 186


RESULT 24
US-09-900-575-33
; Sequence 33, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-33

  Query Match             96.6%;  Score 819;  DB 10;  Length 279;
  Best Local Similarity   95.7%;  Pred. No. 2.2e-80;
  Matches  154;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||:|||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||: |||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||| |||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVY 186


RESULT 25
US-09-900-575-41
; Sequence 41, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 41
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-41

  Query Match             96.6%;  Score 819;  DB 10;  Length 279;
  Best Local Similarity   96.3%;  Pred. No. 2.2e-80;
  Matches  155;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           | ||||||||||||||||||||||||||||||||||||||||:|||||||:|||||||
Db      26 PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           |||||||||||||||||||||||| |||||||: ||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTLVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 26
US-09-732-350-3
; Sequence 3, Application US/09732350
; Patent No. US20010031490A1
;  GENERAL INFORMATION:
;    APPLICANT:  Svendsen, Allan
;    APPLICANT:  Xu, Feng
;    TITLE OF INVENTION:  LACCASE MUTANTS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. US20010031490A1o No. US20010031490A1disk of No. US20010031490A
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/732,350
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/032,315
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rozek, Carol
;      REGISTRATION NUMBER:  36,993
;      REFERENCE/DOCKET NUMBER:  5200.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  499 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-732-350-3

  Query Match              9.8%;  Score 83;  DB 10;  Length 499;
  Best Local Similarity   28.5%;  Pred. No. 0.36;
  Matches   39;  Conservative   19;  Mismatches   49;  Indels   30;  Gaps    8;

Qy       7 QNLVVDLSTQIFCHNDYP------ETITDYVTLQRGSAYGGVLSNFSGTV-----KYSGS 55
           | |:|| | |||    |       :|: :|  ::    :| |   |:| :     :| |:
Db     229 QPLLVD-SIQIFAAQRYSFVLNANQTVGNY-WVRANPNFGTV--GFAGGINSAILRYQGA 284

Qy      56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSD 115
             |||::|| |:    |:   |:|    | |   | | ||       |    |:|
Db     285 PVAEPTTTQTPSVIPLIETNL-HPLARMPVPGSPTPGGVDKA--------LNLAFNFNGT 335
```

```
Qy    116 DFQFVWNIYANNDVVVP 132
          :|      : ||    |
Db    336 NF------FINNATFTP 346


RESULT 27
US-09-815-242-12939
; Sequence 12939, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12939
;   LENGTH: 358
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12939

  Query Match             8.9%;  Score 75.5;  DB 10;  Length 358;
  Best Local Similarity   24.6%;  Pred. No. 1.5;
  Matches   28;  Conservative   19;  Mismatches   44;  Indels   23;  Gaps    4;

Qy     11 VDLSTQIFCHNDYPETIT--DYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRV 68
          ::::|:|   :   | :|  | | ||   | | ||      :| :| :  |   :| |
Db    194 MEMNTRIQVEHPVTEMVTGIDLVKLQLQVAMGDVLPYKQEDIKLTGHAIEFRINAENP-- 251

Qy     69 VYNSRTDKPWPVALYLTPVSSAGGLVIKA-------------GSLIAVLILRQ 108
           | :    |  :  || |    ||  ::                |::| ||: :
Db    252 -YKNFMPSPGKIEQYLAP----GGYGVRIESACYTNYTIPPYYDSMVAKLIIHE 300


RESULT 28
US-09-815-242-5806
; Sequence 5806, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5806
;   LENGTH: 448
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(448)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-5806

  Query Match             8.9%;  Score 75.5;  DB 10;  Length 448;
  Best Local Similarity   24.6%;  Pred. No. 2;
  Matches   28;  Conservative   19;  Mismatches   44;  Indels   23;  Gaps    4;

Qy     11 VDLSTQIFCHNDYPETIT--DYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRV 68
          ::::|:|   :   | :|  | | ||   | | ||      :| :| :  |   :| |
Db    288 MEMNTRIQVEHPVTEMVTGIDLVKLQLQVAMGDVLPYKQEDIKLTGHAIEFRINAENP-- 345

Qy     69 VYNSRTDKPWPVALYLTPVSSAGGLVIKA-------------GSLIAVLILRQ 108
           | :    |  :  || |    ||  ::                |::| ||: :
Db    346 -YKNFMPSPGKIEQYLAP----GGYGVRIESACYTNYTIPPYYDSMVAKLIIHE 394


RESULT 29
US-09-815-242-13127
; Sequence 13127, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
```

```
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13127
;   LENGTH: 451
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-13127

  Query Match             8.9%;  Score 75.5;  DB 10;  Length 451;
  Best Local Similarity   24.6%;  Pred. No. 2;
  Matches   28;  Conservative   19;  Mismatches   44;  Indels   23;  Gaps    4;

Qy     11 VDLSTQIFCHNDYPETIT--DYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRV 68
          ::::|:|   :  | :|  |:| ||   | | ||      :| :| :  |   :| |
Db    287 MEMNTRIQVEHPVTEMVTGIDLVKLQLQVAMGDVLPYKQEDIKLTGHAIEFRINAENP-- 344

Qy     69 VYNSRTDKPWPVALYLTPVSSAGGLVIKA-------------GSLIAVLILRQ 108
           | :    |  :  || |    ||  ::               |::| ||: :
Db    345 -YKNFMPSPGKIEQYLAP----GGYGVRIESACYTNYTIPPYYDSMVAKLIIHE 393


RESULT 30
US-09-804-551B-18
; Sequence 18, Application US/09804551B
; Patent No. US20020056151A1
; GENERAL INFORMATION:
;  APPLICANT: Bayer Aktiengesellschaft
;  TITLE OF INVENTION: Receptors for peptides from insects
;  FILE REFERENCE: Le A 34 394
;  CURRENT APPLICATION NUMBER: US/09/804,551B
;  CURRENT FILING DATE:  2001-03-12
;  PRIOR APPLICATION NUMBER: DE 100 13 618.4
;  PRIOR FILING DATE: 2000-03-18
;  NUMBER OF SEQ ID NOS: 92
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 410
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-804-551B-18

  Query Match             8.4%;  Score 71;  DB 10;  Length 410;
  Best Local Similarity   28.5%;  Pred. No. 5.4;
  Matches   37;  Conservative   12;  Mismatches   37;  Indels   44;  Gaps    7;

Qy      1 PVVNVGQNL-----VVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGS 55
          |: || : |     ||  |||             :  |  ||:      |||  |
Db     23 PMPNVTETLWDLAMVVSQSTQ------------WPLLDTGSS-----ENFSELV----- 59

Qy     56 SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKA---GSLIAV-LILRQTNN 111
             |:||| | |  | :       |: |:  |   |:     |:|| | | :||  |
Db     60 ------TTETPYVPYGRRPE------TYIVPILFALIFVVGVLGNGTLIVVFLSVRQMRN 107

Qy    112 YNSDDFQFVW 121
            :   |  |
Db    108 VPNTGGQRTW 117


RESULT 31
US-09-881-752A-18
; Sequence 18, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
;  APPLICANT: Kleanthous, Harold
;  APPLICANT:  Al-Garawi, Amal
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Tomb, Jean-Francois
;  APPLICANT:  Oomen, Raymond P.
;  TITLE OF INVENTION: Identification of Polynucleotides
;  TITLE OF INVENTION:  Encoding No. US20020115078A1el Helicobacter Polypeptides in the
;  TITLE OF INVENTION:  Genome
;  FILE REFERENCE: 06132/041002
;  CURRENT APPLICATION NUMBER: US/09/881,752A
;  CURRENT FILING DATE:  2001-06-15
;  PRIOR APPLICATION NUMBER: US 08/833,457
;  PRIOR FILING DATE: 1997-04-01
;  NUMBER OF SEQ ID NOS: 370
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
;   LENGTH: 471
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-09-881-752A-18

  Query Match             8.3%;  Score 70.5;  DB 10;  Length 471;
  Best Local Similarity   33.3%;  Pred. No. 7.3;
  Matches   19;  Conservative    8;  Mismatches   27;  Indels    3;  Gaps    1;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSY 57
          |    ||||::    | |  |:|  :|  |   |  : : | : | || :  | | |
Db    312 PATQNGQNLII---LQRFDINNYNVSIAFYKVFQNANGWIGNMGNPSGVIMGSNSVY 365


RESULT 32
US-09-815-242-13697
; Sequence 13697, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13697
;   LENGTH: 511
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-815-242-13697

  Query Match             8.3%;  Score 70.5;  DB 10;  Length 511;
  Best Local Similarity   23.2%;  Pred. No. 8.2;
  Matches   29;  Conservative   14;  Mismatches   43;  Indels   39;  Gaps    4;

Qy      7 QNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVL--------SNFSGTVKYSGSSYP 58
          | |: |||  :|                |  ||  |          ||:| : | |  |
Db    107 QQLINDLSKSLF----------------GGGYGDKLFPPTIQVNPNFTGAISYQGLDYV 149

Qy     59 F------PTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNY 112
                  ::     ||    |  |: |:|           | |:    |::|:  :
Db    150 SLEGEFGQDFAQLAYWAYNIMVQKTLPIELWLE--------YEKEGNCDFRLVIRKMWSG 201
```

```
Qy    113 NSDDF 117
          : |||
Db    202 SVDDF 206


RESULT 33
US-09-741-233A-6
; Sequence 6, Application US/09741233A
; Patent No. US20020157142A1
; GENERAL INFORMATION:
;  APPLICANT: Glover, Lyn W.
;  APPLICANT:  Budziszewski, Gregory J.
;  APPLICANT:  Zhou, Qing
;  APPLICANT:  Levin, Joshua Z.
;  TITLE OF INVENTION: Herbicide Target Genes and Methods
;  FILE REFERENCE: PB/5-31127A
;  CURRENT APPLICATION NUMBER: US/09/741,233A
;  CURRENT FILING DATE:  2000-12-18
;  NUMBER OF SEQ ID NOS: 21
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 184
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-741-233A-6

  Query Match             8.1%;  Score 69;  DB 9;  Length 184;
  Best Local Similarity   28.8%;  Pred. No. 3;
  Matches   30;  Conservative   17;  Mismatches   23;  Indels   34;  Gaps    9;

Qy     19 CHNDYPETITDYVTL-------QRGSAYGGVLSNFSGTVKYSGS--SY---PFP------ 60
          | |  | |:  : :        ::| : |    :|:||  :| |  ||   |||
Db     56 CFN--PNTVKSHCSYAVNSFFQKKGQSLG--TCDFAGTATFSASDPSYTTCPFPASASGS 111

Qy     61 ------TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAG 98
                ||: : ||   : | :|: :    || |: ||| | :|
Db    112 GTTTPVTTTPSTRVPTTTNT-RPYTI----TP-STGGGLGIPSG 149


RESULT 34
US-09-970-638-5
; Sequence 5, Application US/09970638
; Patent No. US20020090627A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Meyers, Rachel
;  TITLE OF INVENTION: 27419, A NOVEL HUMAN ARGININE-N-METHYL
;  TITLE OF INVENTION:  TRANSFERASE AND USES THEREOF
;  FILE REFERENCE: 38155-20039.00
;  CURRENT APPLICATION NUMBER: US/09/970,638
;  CURRENT FILING DATE:  2002-03-12
;  PRIOR APPLICATION NUMBER: US 60/237,717
;  PRIOR FILING DATE: 2000-10-05
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 185
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Consensus amino acid sequence
US-09-970-638-5

  Query Match             8.1%;  Score 69;  DB 10;  Length 185;
  Best Local Similarity   23.1%;  Pred. No. 3;
  Matches   37;  Conservative   21;  Mismatches   58;  Indels   44;  Gaps    7;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSN-------FSGTVKYS 53
          |:|:   | ||| :          : |||  |::    :   :|:        |  ||| :
Db     43 PIVDPKYNHVVDPN----------QIITDPCTIKEFDMHTVKISDLETFTSPFKLTVKRN 92

Qy     54 GSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG-----------------LVI 95
          | . : |    :   | : |:   | | |:   ||  : |                    | :
Db     93 GMLHAFTAWFD---VYFTSKRCHPIPEAID-KPVGFSTGPQSPYTHWKQTIFYMEDPLPV 148

Qy     96 KAGSLI--AVLILRQTNNYNSDDFQFVWNIYANNDVVVPT 133
          | |  |  :: | :   |    |    ||:    |   ||
Db    149 KTGDTIEGSITISKNKKNNRDLDITLSWNVNGRQD---PT 185


RESULT 35
US-09-995-598-24
; Sequence 24, Application US/09995598
; Patent No. US20020142293A1
; GENERAL INFORMATION:
;  APPLICANT: CRAINIC, RADU
;  APPLICANT:  CARO, VALERIE
;  APPLICANT:  GUILLOT, SOPHIE
;  TITLE OF INVENTION: ENTEROVIRAL POLYNUCLEOTIDES, METHODS OF DETECTING ENTEROVIRUSE
;  TITLE OF INVENTION:  CONTAINING THE POLYNUCLEOTIDES
;  FILE REFERENCE: 215059US-660-660-0
;  CURRENT APPLICATION NUMBER: US/09/995,598
;  CURRENT FILING DATE:  2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/253741
;  PRIOR FILING DATE: 2000-11-29
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 245
;   TYPE: PRT
;   ORGANISM: Echovirus
US-09-995-598-24

  Query Match             8.1%;  Score 69;  DB 10;  Length 245;
  Best Local Similarity   21.4%;  Pred. No. 4.4;
  Matches   36;  Conservative   22;  Mismatches   68;  Indels   42;  Gaps    9;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          | :::|              | |      :    :   ||    |  | : :   : | |
Db     16 PFISIG-------------NAYSNFYDGWSHFSQNGVYGFTTLNNMGQLFFRHVNKPNP 61

Qy     61 TT-SETPRVVYNSRTDKPW---PVAL--YLT---------PVSSAGGLVIKAGSL----- 100
           | :   |: :  :  : |   |  |  |:          ||:     :|  |:
Db     62 ATITSVARIYFKPKHVRAWVPRPPRLCPYINSGNVNFDPKPVTEVRSSIITTGAFGQQSG 121

Qy    101 -IAVLILRQTNNY---NSDDFQFVWNIYANNDVVVPT---GGCDVSAR 141
           : |   |  | :    :|    ||  | | |::| |    |||  ||
Db    122 AVYVGNYRVVNRHLATRNDWLSCVWEDY-NRDLLVSTTTAHGCDTIAR 168


RESULT 36
US-09-841-132-309
; Sequence 309, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
;  APPLICANT: Bhatia, Ajay
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Probst, Peter
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
;  TITLE OF INVENTION:  DIAGNOSIS OF CHLAMYDIAL INFECTION
;  FILE REFERENCE: 210121.469C8
;  CURRENT APPLICATION NUMBER: US/09/841,132
;  CURRENT FILING DATE:  2001-04-23
;  NUMBER OF SEQ ID NOS: 599
;  SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 309
;   LENGTH: 619
;   TYPE: PRT
;   ORGANISM: Chlamydia trachomatis
US-09-841-132-309

  Query Match             8.1%;  Score 69;  DB 10;  Length 619;
  Best Local Similarity   21.3%;  Pred. No. 15;
  Matches   27;  Conservative   21;  Mismatches   61;  Indels   18;  Gaps    5;
```

```
Qy      31 VTLQRGSA-YGGVLSNFSGTVKYSGSSYPFP----TTSETPRVVYNSR---TDKPWPVAL 82
           :| :  || :|| : | :| |:: |:: |      ||      :| |      :  |:
Db     402 ITFEGNSAVHGGAIYNKNGLVEFLGNAGPLAFKENTTIANGGAIYTSNFKANQQTSPILF 461

Qy      83 YLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD 142
              :  ||         |:         | |  :|  |        : :  | ::| :
Db     462 SQNHANKKGG---------AIYAQYVNLEQNQDTIRFEKNTAKEGGGAITSSQCSITAHN 512

Qy     143 VTVTLPD 149
           |:|  |
Db     513 -TITFSD 518


RESULT 37
US-09-864-761-37319
; Sequence 37319, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37319
;   LENGTH: 5701
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC010680.3
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.95
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.8
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =   1
;   OTHER INFORMATION: SWISSPROT HIT: P16419, EVALUE 2.00e-73
;   OTHER INFORMATION: EST_HUMAN HIT: AW384459.1, EVALUE 5.00e-90
;   OTHER INFORMATION: EST_HUMAN HIT: AW384459.1, EVALUE 5.00e-90
US-09-864-761-37319

  Query Match             8.1%;  Score 69;  DB 10;  Length 5701;
  Best Local Similarity   25.0%;  Pred. No. 3e+02;
  Matches   36;  Conservative   16;  Mismatches   58;  Indels   34;  Gaps    8;

Qy      31 VTLQRGSAY--GGVLSNFSGTVKYSGSS-----YPF--PTTSETPRVVYNSRTDK--PWP 79
           | |  || |       |  |   || ||     |||  |.   ||:||: :::     |
Db    4878 VRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYPFSPPGPPGTPKVVHATKSTMLVTWQ 4937

Qy      80 VALYLTPVSSAGGLVI--------------KAGSLIAVLILRQTNNYNSDDFQFVWNIY 124
           |    ||:  |  ||               ||  ::      : :  : :   : : :|
Db    4938 V-----PVNDGGSRVIGYHLEYKERSSILWSKANKILIADTQMKVSGLD-EGLMYEYRVY 4991

Qy     125 ANNDVVVPTGGCDVSARDVTVTLP 148
           | |  :   | | |   |    |
Db    4992 AEN--IAGIGKCSKSCEPVPARDP 5013


RESULT 38
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12610
;   LENGTH: 5795
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12610

  Query Match             8.1%;  Score 69;  DB 10;  Length 5795;
  Best Local Similarity   19.2%;  Pred. No. 3.1e+02;
  Matches   35;  Conservative   23;  Mismatches   58;  Indels   66;  Gaps    7;

Qy      16 QIFCHN-------DYP----ETITDYVTLQRGSAYG------------------GVLSN 45
           :|  |:       |:|     :  ||  |  |||                    |   |
Db    1679 EIVAHSHYTVQGQDFPAGNGSSAADYFKLSNGSAIPDATITWVSGQAPNKDNTRIGEDIN 1738
```

```
Qy      46 FSGTVKYSGSSYPFPTT------SETPRVVYNSRTDKPWP-------VALYLTPVSSAGG 92
           :  :   | : |   |      |  |: |: :     :|          |: ||: :
Db    1739 VTANILIDGETTPITKTATYKVVSSVPKHVFETNRGAVFPGVSDVYDAKQYVKPVNDS-- 1796

Qy      93 LVIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRG 152
                        | |    :|||  :   : |||       :| ||: ||  :::
Db    1797 --------------WTQNAQRMNFQFTNSYGPSKDVV------GISTRDIRVTYDNHQT 1835

Qy     153 SV 154
           :
Db    1836 QI 1837


RESULT 39
US-09-995-598-50
; Sequence 50, Application US/09995598
; Patent No. US20020142293A1
; GENERAL INFORMATION:
;  APPLICANT: CRAINIC, RADU
;  APPLICANT:  CARO, VALERIE
;  APPLICANT:  GUILLOT, SOPHIE
;  TITLE OF INVENTION: ENTEROVIRAL POLYNUCLEOTIDES, METHODS OF DETECTING ENTEROVIRUSES A
;  TITLE OF INVENTION:  CONTAINING THE POLYNUCLEOTIDES
;  FILE REFERENCE: 215059US-660-660-0
;  CURRENT APPLICATION NUMBER: US/09/995,598
;  CURRENT FILING DATE:  2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/253741
;  PRIOR FILING DATE: 2000-11-29
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
;   LENGTH: 239
;   TYPE: PRT
;   ORGANISM: Echovirus
US-09-995-598-50

  Query Match             8.1%;  Score 68.5;  DB 10;  Length 239;
  Best Local Similarity   26.1%;  Pred. No. 4.8;
  Matches   42;  Conservative   16;  Mismatches   46;  Indels   57;  Gaps   11;

Qy      36 GSAYGGVL---SNFS-------GTVKYSGSSY---------PFPTTSETPRVVYNSRTDKP 77
           |:||       |:||        |: : |  |         | | || | || :  :  |
Db      18 GNAYSNFYDGWSHFSQNGVYGYNTLNHMGQLYMRHVNGPSPLPMTS-TVRVYFKPKHVKA 76

Qy      78 W---PVAL-----YLTPVSSAGGLVIKAGSL--IAVLILRQTNNYNS------------- 114
           |   |  |       |   |:  :  | ||:  |   :    :|| :
Db      77 WVPRPPRLCQYKNASTVNFSSTNITDKRGSITYIPDTVKPDVSNYGAFGQQSGAVYVGNY 136

Qy     115 ----------DDFQ-FVWNIYANNDVVVPT---GGCDVSAR 141
                    :|:|  ||  | | |::| |    |||  ||
Db     137 RVVNRHLATYNDWQNCVWEDY-NRDLLVSTTTAHGCDTIAR 176


RESULT 40
US-09-995-598-12
; Sequence 12, Application US/09995598
; Patent No. US20020142293A1
; GENERAL INFORMATION:
;  APPLICANT: CRAINIC, RADU
;  APPLICANT:  CARO, VALERIE
;  APPLICANT:  GUILLOT, SOPHIE
;  TITLE OF INVENTION: ENTEROVIRAL POLYNUCLEOTIDES, METHODS OF DETECTING ENTEROVIRUSES A
;  TITLE OF INVENTION:  CONTAINING THE POLYNUCLEOTIDES
;  FILE REFERENCE: 215059US-660-660-0
;  CURRENT APPLICATION NUMBER: US/09/995,598
;  CURRENT FILING DATE:  2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/253741
;  PRIOR FILING DATE: 2000-11-29
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 244
;   TYPE: PRT
;   ORGANISM: Echovirus
US-09-995-598-12

  Query Match             8.1%;  Score 68.5;  DB 10;  Length 244;
  Best Local Similarity   22.7%;  Pred. No. 5;
  Matches   41;  Conservative   21;  Mismatches   62;  Indels   57;  Gaps    9;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGT--VKYSGSSYP 58
           | :::|              | |      :    :   ||    |  |   |::   | |
Db      13 PFISIG-------------NAYSNFYDGWSHFSQNGVYGYNTLNHMGQIYVRHLNGSSP 58

Qy      59 FPTTSETPRVVYNSRTDKPW-----------PVALYLTP------------------- 86
             | || | |: :  :  | |               :  ||
Db      59 LPMTS-TVRMYFKPKHVKVWVPRPPRLCQYKNASTVNFTPTNITEKRKSINYIPETVKPD 117

Qy      87 VSSAGGLVIKAGSLIAVLILRQTNNY---NSDDFQFVWNIYANNDVVVPT---GGCDVSA 140
           ||: |    ::|: | |   |  | :   ::|    ||: | | |::| |    |||  |
Db     118 VSTYGAFGQQSGA-IYVGNYRVVNRHLATHTDWQNCVWDNY-NRDLLVSTTTAHGCDTIA 175

Qy     141 R 141
           |
Db     176 R 176


RESULT 41
US-09-905-291A-104
; Sequence 104, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
;  APPLICANT: Genentech, Inc.
;  APPLICANT:  Ashkenazi, Avi
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan L.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, A.
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth, J.
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Mather, Jennie P.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William, I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: 10466-14
;  CURRENT APPLICATION NUMBER: US/09/905,291A
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: PCT/US00/04414
;  PRIOR FILING DATE: 2000-02-22
;  PRIOR APPLICATION NUMBER: US 60/143,048
;  PRIOR FILING DATE: 1999-07-07
;  PRIOR APPLICATION NUMBER: US 60/145,698
;  PRIOR FILING DATE: 1999-07-26
;  PRIOR APPLICATION NUMBER: US 60/146,222
;  PRIOR FILING DATE: 1999-07-28
;  PRIOR APPLICATION NUMBER: PCT/US99/20594
;  PRIOR FILING DATE: 1999-09-08
;  PRIOR APPLICATION NUMBER: PCT/US99/20944
;  PRIOR FILING DATE: 1999-09-13
;  PRIOR APPLICATION NUMBER: PCT/US99/21090
```

```
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/21547
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/23089
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: PCT/US99/28214
;  PRIOR FILING DATE: 1999-11-29
;  PRIOR APPLICATION NUMBER: PCT/US99/28313
;  PRIOR FILING DATE: 1999-11-30
;  PRIOR APPLICATION NUMBER: PCT/US99/28564
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/28565
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/30095
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: PCT/US99/30911
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US99/30999
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US00/00219
;  PRIOR FILING DATE: 2000-01-05
;  NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 104
;   LENGTH: 415
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-905-291A-104

  Query Match             8.1%;  Score 68.5;  DB 9;  Length 415;
  Best Local Similarity   19.7%;  Pred. No. 10;
  Matches   35;  Conservative   20;  Mismatches   56;  Indels   67;  Gaps    6;

Qy      12 DLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYN 71
           |:    :|  ||                 || :::     || | | | |  ||   ::
Db     201 DVERDNYCRYDYVAVFN----------GGEVNDARRIGKYCGDSPPAPIVSERNELLIQ 249

Qy      72 ------------------------SRTDKP----WPVALYLTP-VSSAGGLVIKAGSLIA 102
                                    : |::|    :||   | | |:       : |:|
Db     250 FLSDLSLTADGFIGHYIFRPKKLPTTTEQPVTTTFPVTTGLKPTVALCQQKCRRTGTL-- 307

Qy     103 VLILRQTNNYNSDDFQF------------------VWNIYANNDVVVPTGGCDVSAR 141
               || | ||                      : |||   :: :   | ::|||
Db     308 ------EGNYCSSDFVLAGTVITTITRDGSLHATVSIINIYKEGNLAIQQAGKNMSAR 359


RESULT 42
US-09-909-320-104
; Sequence 104, Application US/09909320
; Patent No. US20020132240A1
; GENERAL INFORMATION:
;  APPLICANT: Genentech, Inc.
;  APPLICANT:  Ashkenazi, Avi
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan L.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, A.
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth, J.
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Mather, Jennie P.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William, I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: 10466-14
;  CURRENT APPLICATION NUMBER: US/09/909,320
;  CURRENT FILING DATE:  2002-01-04
;  PRIOR APPLICATION NUMBER: PCT/US00/04414
;  PRIOR FILING DATE: 2000-02-22
;  PRIOR APPLICATION NUMBER: US 60/143,048
;  PRIOR FILING DATE: 1999-07-07
;  PRIOR APPLICATION NUMBER: US 60/145,698
;  PRIOR FILING DATE: 1999-07-26
;  PRIOR APPLICATION NUMBER: US 60/146,222
;  PRIOR FILING DATE: 1999-07-28
;  PRIOR APPLICATION NUMBER: PCT/US99/20594
;  PRIOR FILING DATE: 1999-09-08
;  PRIOR APPLICATION NUMBER: PCT/US99/20944
;  PRIOR FILING DATE: 1999-09-13
;  PRIOR APPLICATION NUMBER: PCT/US99/21090
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/21547
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/23089
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: PCT/US99/28214
;  PRIOR FILING DATE: 1999-11-29
;  PRIOR APPLICATION NUMBER: PCT/US99/28313
;  PRIOR FILING DATE: 1999-11-30
;  PRIOR APPLICATION NUMBER: PCT/US99/28564
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/28565
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/30095
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: PCT/US99/30911
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US99/30999
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US00/00219
;  PRIOR FILING DATE: 2000-01-05
;  NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 104
;   LENGTH: 415
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-909-320-104

  Query Match             8.1%;  Score 68.5;  DB 10;  Length 415;
  Best Local Similarity   19.7%;  Pred. No. 10;
  Matches   35;  Conservative   20;  Mismatches   56;  Indels   67;  Gaps    6;

Qy      12 DLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYN 71
           |:    :|  ||                 || :::     || | | | |  ||   ::
Db     201 DVERDNYCRYDYVAVFN----------GGEVNDARRIGKYCGDSPPAPIVSERNELLIQ 249

Qy      72 ------------------------SRTDKP----WPVALYLTP-VSSAGGLVIKAGSLIA 102
                                    : |::|    :||   | | |:       : |:|
Db     250 FLSDLSLTADGFIGHYIFRPKKLPTTTEQPVTTTFPVTTGLKPTVALCQQKCRRTGTL-- 307

Qy     103 VLILRQTNNYNSDDFQF------------------VWNIYANNDVVVPTGGCDVSAR 141
               || | ||                      : |||   :: :   | ::|||
Db     308 ------EGNYCSSDFVLAGTVITTITRDGSLHATVSIINIYKEGNLAIQQAGKNMSAR 359


RESULT 43
US-09-909-088B-104
; Sequence 104, Application US/09909088B
; Patent No. US20020146709A1
; GENERAL INFORMATION:
;  APPLICANT: Genentech, Inc.
;  APPLICANT:  Ashkenazi, Avi
;  APPLICANT:  Botstein, David
```

```
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan L.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, A.
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth, J.
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Mather, Jennie P.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William, I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: 10466-14
;  CURRENT APPLICATION NUMBER: US/09/909,088B
;  CURRENT FILING DATE:  2001-07-18
;  PRIOR APPLICATION NUMBER: PCT/US00/04414
;  PRIOR FILING DATE: 2000-02-22
;  PRIOR APPLICATION NUMBER: US 60/143,048
;  PRIOR FILING DATE: 1999-07-07
;  PRIOR APPLICATION NUMBER: US 60/145,698
;  PRIOR FILING DATE: 1999-07-26
;  PRIOR APPLICATION NUMBER: US 60/146,222
;  PRIOR FILING DATE: 1999-07-28
;  PRIOR APPLICATION NUMBER: PCT/US99/20594
;  PRIOR FILING DATE: 1999-09-08
;  PRIOR APPLICATION NUMBER: PCT/US99/20944
;  PRIOR FILING DATE: 1999-09-13
;  PRIOR APPLICATION NUMBER: PCT/US99/21090
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/21547
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/23089
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: PCT/US99/28214
;  PRIOR FILING DATE: 1999-11-29
;  PRIOR APPLICATION NUMBER: PCT/US99/28313
;  PRIOR FILING DATE: 1999-11-30
;  PRIOR APPLICATION NUMBER: PCT/US99/28564
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/28565
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/30095
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: PCT/US99/30911
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US99/30999
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US00/00219
;  PRIOR FILING DATE: 2000-01-05
;  NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 104
;   LENGTH: 415
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-909-088B-104

  Query Match             8.1%;  Score 68.5;  DB 10;  Length 415;
  Best Local Similarity   19.7%;  Pred. No. 10;
  Matches   35;  Conservative   20;  Mismatches   56;  Indels   67;  Gaps    6;

Qy      12 DLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYN 71
```

```
           |:    :|  ||                 || :::     || | | | | |  ||   ::
Db     201 DVERDNYCRYDYVAVFN----------GGEVNDARRIGKYCGDSPPAPIVSERNELLIQ 249

Qy      72 ----------------------SRTDKP----WPVALYLTP-VSSAGGLVIKAGSLIA 102
                                : |::|    :||   | | |:       : |:|
Db     250 FLSDLSLTADGFIGHYIFRPKKLPTTTEQPVTTTFPVTTGLKPTVALCQQKCRRTGTL-- 307

Qy     103 VLILRQTNNYNSDDFQF-----------------VWNIYANNDVVVPTGGCDVSAR 141
                 || | ||                     : |||    :: :   | ::|||
Db     308 ------EGNYCSSDFVLAGTVITTITRDGSLHATVSIINIYKEGNLAIQQAGKNMSAR 359


RESULT 44
US-09-886-468-19
; Sequence 19, Application US/09886468
; Patent No. US20020037293A1
; GENERAL INFORMATION:
;  APPLICANT: Aventis Pasteur Limited
;  TITLE OF INVENTION: Chlamydia antigens and corresponding DNA fragments and uses th
;  FILE REFERENCE: 77813-5
;  CURRENT APPLICATION NUMBER: US/09/886,468
;  CURRENT FILING DATE:  1999-12-23
;  PRIOR APPLICATION NUMBER: 60/113,280
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/113,281
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/113,282
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/113,283
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/113,284
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/113,285
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/113,385
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/114,050
;  PRIOR FILING DATE: 1998-12-28
;  PRIOR APPLICATION NUMBER: 60/114,056
;  PRIOR FILING DATE: 1998-12-28
;  PRIOR APPLICATION NUMBER: 60/114,057
;  PRIOR FILING DATE: 1998-12-28
;  PRIOR APPLICATION NUMBER: 60/114,058
;  PRIOR FILING DATE: 1998-12-28
;  PRIOR APPLICATION NUMBER: 60/114,059
;  PRIOR FILING DATE: 1998-12-28
;  PRIOR APPLICATION NUMBER: 60/114,061
;  PRIOR FILING DATE: 1998-12-28
;  NUMBER OF SEQ ID NOS: 26
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 922
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
US-09-886-468-19

  Query Match             8.1%;  Score 68.5;  DB 10;  Length 922;
  Best Local Similarity   27.4%;  Pred. No. 30;
  Matches   31;  Conservative   15;  Mismatches   34;  Indels   33;  Gaps    6;

Qy      53 SGSSYPFP--TTSETPRVVYNSRTDKPWPVALYLTPVS-SAGGLVIKAGSLIAVLILRQT 109
           |||| | |  | |:  ::|:          |:|   :| |:||  :
Db     264 SGSSTPVPIVTFSDNKQLVFERNHSIMGGGAIYARKLSISSGGPTL------------- 309

Qy     110 NNYNSDDFQFVWNI-YANND-----VVVPTGG-CDVSARDVTVTLPDYRGSVP 155
                  |: || |||:      : : |||   :||   |:|    | |:|
Db     310 ---------FINNISYANSQNLGGAIAIDTGGEISLSAEKGTITFQGNRTSLP 353


RESULT 45
US-09-860-846-33
; Sequence 33, Application US/09860846
; Patent No. US20020164742A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 33
;   LENGTH: 3739
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-33

  Query Match             8.0%;  Score 68;  DB 9;  Length 3739;
  Best Local Similarity   23.4%;  Pred. No. 2.2e+02;
  Matches   41;  Conservative   25;  Mismatches   61;  Indels   48;  Gaps    8;

Qy      24 PETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPF----PTTSETPRV-VYNSRTDKPW 78
           |||:|   ||:| :     |:       :| :  :    |||:  | :  |  :|:: |
Db    2348 PETVTGLGTLRRDNGGQHRLTTSLAEAWANGLTVDWASLLPTTTTHPDLPTYAFQTERYW 2407

Qy      79 PVALYLTPVSSAGGLVIKA----------------GSLIAVLILRQTNNYNSD--- 115
           |      |  || | :  |                | |:   :  :|: : :|
Db    2408 P-----QPDLSAAGDITSAGLGAAEHPLLGAAVALADSDGCLLTGSLSLRTHPWLADHAV 2462

Qy     116 -------DFQFVWNIYANNDVVVPTGGCDVSAR---DVTVTLPDYRGSVPIPLTV 160
                  ||   :   | |    |||:      |  : ||  ||:| : |:|
Db    2463 AGTVLLPGTAFVELAFRAGDQV----GCDLVEELTLDAPLVLP-RRGAVRVQLSV 2512


Search completed: November 28, 2002, 19:01:24
Job time : 23 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 18:54:01 ; Search time 66 Seconds
(without alignments)
325.051 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 161
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	161	100.0	279	23	AAE18419	Escherichia coli s
2	92	57.1	268	23	ABG68154	Escherichia coli m
3	92	57.1	279	22	AAB47074	Adhesin protein, F
4	92	57.1	279	22	AAB72839	Bacterial type 1 p
5	92	57.1	279	22	AAY72515	Escherichia coli F
6	92	57.1	279	23	AAE18424	Escherichia coli s
7	92	57.1	279	23	AAE18426	Escherichia coli s
8	92	57.1	279	23	AAE18433	Escherichia coli s
9	92	57.1	296	22	AAB47073	Immunogenic comple
10	92	57.1	300	16	AAR76745	FimH protein deriv
11	92	57.1	300	21	AAB16009	E. coli proliferat
12	92	57.1	300	21	AAY59456	E. coli PC31 FimH
13	92	57.1	300	23	AAU77488	Escherichia coli t
14	92	57.1	304	22	AAB47072	Immunogenic comple
15	90	55.9	279	23	AAE18417	Escherichia coli s
16	90	55.9	279	23	AAE18418	Escherichia coli s
17	90	55.9	279	23	AAE18422	Escherichia coli s
18	90	55.9	279	23	AAE18429	Escherichia coli s
19	90	55.9	279	23	AAE18432	Escherichia coli s
20	90	55.9	279	23	AAE18436	Escherichia coli F
21	90	55.9	300	16	AAR76769	FimH protein deriv
22	90	55.9	300	16	AAR76771	FimH protein deriv
23	90	55.9	300	16	AAR76773	FimH protein deriv
24	90	55.9	300	16	AAR76774	FimH protein deriv
25	90	55.9	300	16	AAR76763	FimH protein deriv
26	89	55.3	296	16	AAR76767	FimH protein deriv
27	79	49.1	300	16	AAR76772	FimH protein deriv
28	75	46.6	280	23	AAE18420	Escherichia coli s
29	67	41.6	279	23	AAE18415	Escherichia coli s
30	67	41.6	279	23	AAE18416	Escherichia coli s
31	67	41.6	279	23	AAE18425	Escherichia coli s
32	67	41.6	279	23	AAE18427	Escherichia coli s
33	67	41.6	279	23	AAE18430	Escherichia coli s
34	67	41.6	279	23	AAE18431	Escherichia coli s
35	67	41.6	279	23	AAE18434	Escherichia coli s
36	67	41.6	279	23	AAE18435	Escherichia coli s
37	60	37.3	279	23	AAE18428	Escherichia coli s
38	59	36.6	300	16	AAR76768	FimH protein deriv
39	57	35.4	300	16	AAR76766	FimH protein deriv
40	56	34.8	279	23	AAE18423	Escherichia coli s
41	56	34.8	300	16	AAR76765	FimH protein deriv
42	56	34.8	300	16	AAR76770	FimH protein deriv
43	56	34.8	408	23	ABB09458	Fusion protein prt
44	48	29.8	279	23	AAE18413	Escherichia coli s
45	46	28.6	300	16	AAR76764	FimH protein deriv
46	43	26.7	279	23	AAE18414	Escherichia coli s
47	43	26.7	279	23	AAE18421	Escherichia coli s
48	43	26.7	300	16	AAR76775	FimH protein deriv
49	43	26.7	300	16	AAR76776	FimH protein deriv
50	38	23.6	131	21	AAY44760	E. coli type I pil
51	38	23.6	227	21	AAY44761	E. coli type I pil
52	35	21.7	35	22	AAY72511	Escherichia coli F
53	32	19.9	94	22	AAY72513	Novel protein comp
54	31	19.3	31	22	AAY72510	Collagen binding d
55	31	19.3	110	22	AAY72514	Novel protein comp
56	22	13.7	22	22	AAY72509	Mannose binding do
57	22	13.7	71	22	AAY72512	Novel protein comp
58	21	13.0	21	22	AAY72518	MBD-3-ETIT peptide
59	21	13.0	21	22	AAY72520	CHAP-VVVP peptide
60	19	11.8	19	22	AAY72519	CBD-AVLI peptide d
61	16	9.9	18	23	ABG68161	Optimised FimH adh
62	16	9.9	39	23	ABG68238	Optimised FimH adh
63	15	9.3	21	23	ABG68241	Optimised FimH adh
64	13	8.1	20	22	AAY72517	MBD-2-TQIF peptide
65	9	5.6	9	22	AAY72508	Mannose binding do
66	9	5.6	11	23	ABG68160	Optimised FimH adh
67	9	5.6	12	23	ABG68240	Optimised FimH adh
68	9	5.6	32	23	ABG68235	Optimised FimH adh
69	9	5.6	32	23	ABG68236	Optimised FimH adh
70	9	5.6	270	23	AAU91291	Human NOV6 protein
71	9	5.6	275	22	AAE08592	Human NOV13 protei
72	9	5.6	285	22	AAE08590	Human NOV11 protei
73	9	5.6	290	22	AAE08587	Human NOV8 protein
74	8	5.0	39	23	ABG68237	Optimised FimH adh
75	8	5.0	719	22	AAB85513	Human protein kina
76	8	5.0	1021	23	ABP61000	Novel human protei
77	8	5.0	1036	23	ABB80923	Novel human protei
78	7	4.3	9	23	ABG68159	Optimised FimH adh
79	7	4.3	30	23	ABG68233	Optimised FimH adh
80	7	4.3	30	23	ABG68234	Optimised FimH adh
81	7	4.3	96	21	AAG16565	Arabidopsis thalia
82	7	4.3	96	21	AAG48964	Arabidopsis thalia
83	7	4.3	100	21	AAG16564	Arabidopsis thalia

84	7	4.3	100	21	AAG48963	Arabidopsis thalia
85	7	4.3	114	21	AAB40504	Human ORFX ORF268
86	7	4.3	114	23	ABP33123	Human ORF2096 prot
87	7	4.3	117	21	AAG03847	Human secreted pro
88	7	4.3	119	21	AAG10679	Arabidopsis thalia
89	7	4.3	119	21	AAG34783	Arabidopsis thalia
90	7	4.3	120	21	AAG24563	Arabidopsis thalia
91	7	4.3	120	21	AAG43216	Arabidopsis thalia
92	7	4.3	120	21	AAG52566	Arabidopsis thalia
93	7	4.3	129	22	AAO01359	Human polypeptide
94	7	4.3	132	21	AAG12012	Arabidopsis thalia
95	7	4.3	135	21	AAG43215	Arabidopsis thalia
96	7	4.3	136	19	AAW40227	Human myelin P2 pr
97	7	4.3	136	21	AAG24561	Arabidopsis thalia
98	7	4.3	149	21	AAG24388	Arabidopsis thalia
99	7	4.3	149	21	AAG51576	Arabidopsis thalia
100	7	4.3	149	21	AAG51582	Arabidopsis thalia
101	7	4.3	160	21	AAG40839	Zea mays protein f
102	7	4.3	185	23	ABP29807	Streptococcus poly
103	7	4.3	185	23	ABP30311	Streptococcus poly
104	7	4.3	193	23	ABP27453	Streptococcus poly
105	7	4.3	197	21	AAB25172	Eucalyptus grandis
106	7	4.3	210	23	ABB54109	Lactococcus lactis
107	7	4.3	245	21	AAG23051	Arabidopsis thalia
108	7	4.3	245	21	AAG40528	Arabidopsis thalia
109	7	4.3	260	21	AAG51581	Arabidopsis thalia
110	7	4.3	260	23	ABB92475	Herbicidally activ
111	7	4.3	261	21	AAG24387	Arabidopsis thalia
112	7	4.3	263	22	AAG90543	C glutamicum prote
113	7	4.3	274	21	AAG51580	Arabidopsis thalia
114	7	4.3	276	21	AAG24386	Arabidopsis thalia
115	7	4.3	283	21	AAG23050	Arabidopsis thalia
116	7	4.3	283	21	AAG40527	Arabidopsis thalia
117	7	4.3	286	21	AAG51575	Arabidopsis thalia
118	7	4.3	304	21	AAG23049	Arabidopsis thalia
119	7	4.3	304	21	AAG40526	Arabidopsis thalia
120	7	4.3	304	22	AAB96280	Putative P. abyssi
121	7	4.3	349	21	AAB16606	Bacteriophage 192
122	7	4.3	397	22	ABB67545	Drosophila melanog
123	7	4.3	443	23	ABB55143	Lactococcus lactis
124	7	4.3	628	15	AAR54872	Human H13 viral re
125	7	4.3	628	15	AAR54876	Human H13 viral re
126	7	4.3	629	13	AAR25072	H13 gene product.
127	7	4.3	629	15	AAR54868	Full-length human
128	7	4.3	629	15	AAR54869	Human H13 viral re
129	7	4.3	629	15	AAR54870	Human H13 viral re
130	7	4.3	629	15	AAR54871	Human H13 viral re
131	7	4.3	629	15	AAR54873	Human H13 viral re
132	7	4.3	629	15	AAR54874	Human H13 viral re
133	7	4.3	629	15	AAR54875	Human H13 viral re
134	7	4.3	629	15	AAR47121	Full-length human
135	7	4.3	629	20	AAW67473	Human retroviral r
136	7	4.3	707	20	AAW93496	N. meningitidis st
137	7	4.3	722	20	AAW93493	N. meningitidis st
138	7	4.3	772	22	AAU44927	Propionibacterium
139	7	4.3	997	21	AAG40314	Arabidopsis thalia
140	7	4.3	1000	21	AAG31752	Arabidopsis thalia
141	7	4.3	1001	21	AAG40313	Arabidopsis thalia
142	7	4.3	1004	21	AAG31751	Arabidopsis thalia
143	7	4.3	1011	21	AAG40489	Arabidopsis thalia
144	7	4.3	1015	21	AAG40488	Arabidopsis thalia
145	7	4.3	1040	21	AAB15925	E. coli proliferat
146	7	4.3	1040	22	AAU34610	E. coli cellular p
147	7	4.3	1040	22	AAG98972	E. coli growth and
148	7	4.3	1061	22	ABB64235	Drosophila melanog
149	7	4.3	3263	22	ABB67210	Drosophila melanog
150	6	3.7	24	21	AAY15144	Leader peptide of
151	6	3.7	31	21	AAB44979	Human secreted pro
152	6	3.7	31	21	AAB44980	Human secreted pro
153	6	3.7	31	22	AAG75729	Human colon cancer
154	6	3.7	40	22	ABB31517	Peptide #4168 enco
155	6	3.7	40	22	ABB36729	Peptide #4235 enco
156	6	3.7	40	22	ABB22060	Protein #4059 enco
157	6	3.7	40	22	AAM57484	Human brain expres
158	6	3.7	40	22	AAM69879	Human bone marrow
159	6	3.7	40	22	AAM17705	Peptide #4139 enco
160	6	3.7	40	22	AAM30219	Peptide #4256 enco
161	6	3.7	40	22	AAM05367	Peptide #4049 enco
162	6	3.7	40	23	ABG39509	Human peptide enco
163	6	3.7	43	22	ABG13080	Novel human diagno
164	6	3.7	43	22	ABG20394	Novel human diagno
165	6	3.7	46	22	AAB65012	Human secreted pro
166	6	3.7	47	21	AAB34623	Human secreted pro
167	6	3.7	50	22	AAB35585	Antiangiogenic pen
168	6	3.7	52	22	AAU62515	Propionibacterium
169	6	3.7	54	22	ABB28377	Peptide #1028 enco
170	6	3.7	54	22	ABB33555	Peptide #1061 enco
171	6	3.7	54	22	ABB19012	Protein #1011 enco
172	6	3.7	54	22	AAM54329	Human brain expres
173	6	3.7	54	22	AAM66729	Human bone marrow
174	6	3.7	54	22	AAM82460	Human immune/haema
175	6	3.7	54	22	AAM14595	Peptide #1029 enco
176	6	3.7	54	22	AAM27016	Peptide #1053 enco
177	6	3.7	54	22	AAM02320	Peptide #1002 enco
178	6	3.7	54	23	ABG36383	Human peptide enco
179	6	3.7	57	22	AAU68010	Propionibacterium
180	6	3.7	60	21	AAG25494	Arabidopsis thalia
181	6	3.7	60	21	AAY15151	Plasmin Inhibitor
182	6	3.7	60	22	AAU22804	Human prostate can
183	6	3.7	60	22	AAM96110	Human reproductive
184	6	3.7	61	21	AAG03117	Human secreted pro
185	6	3.7	61	22	AAB51020	Early 3/6.7K prote
186	6	3.7	62	22	ABB29727	Peptide #2378 enco
187	6	3.7	62	22	ABB34899	Peptide #2405 enco
188	6	3.7	62	22	ABB20316	Protein #2315 enco
189	6	3.7	62	22	AAU22679	Novel human colon
190	6	3.7	62	22	AAM55709	Human brain expres
191	6	3.7	62	22	AAM68087	Human bone marrow
192	6	3.7	62	22	AAM93125	Human digestive sy
193	6	3.7	62	22	AAM15911	Peptide #2345 enco
194	6	3.7	62	22	AAM28416	Peptide #2453 enco
195	6	3.7	62	22	AAM03647	Peptide #2329 enco
196	6	3.7	62	23	ABG37614	Human peptide enco
197	6	3.7	69	23	ABP03120	Human ORFX protein
198	6	3.7	71	21	AAG25493	Arabidopsis thalia
199	6	3.7	71	22	AAU45137	Propionibacterium
200	6	3.7	73	22	ABG03076	Novel human diagno
201	6	3.7	75	23	AAU75822	Rice L endoxylanas
202	6	3.7	78	22	AAU21275	Human novel foetal
203	6	3.7	78	23	ABP05923	Human ORFX protein
204	6	3.7	81	22	ABB10553	Human pancreatic c
205	6	3.7	81	22	AAM92089	Human digestive sy
206	6	3.7	82	20	AAY07954	Human secreted pro
207	6	3.7	82	21	AAG42148	Zea mays protein f
208	6	3.7	83	21	AAY15145	Proform of Txln 1
209	6	3.7	83	21	AAY15146	Proform of Txln 2
210	6	3.7	83	21	AAY15147	Proform of Txln 3
211	6	3.7	83	21	AAY15148	Proform of Txln 4
212	6	3.7	83	21	AAY15149	Proform of Txln 5
213	6	3.7	83	21	AAY15150	Proform of Txln 6
214	6	3.7	83	22	AAU53165	Propionibacterium
215	6	3.7	83	23	ABP05774	Human ORFX protein
216	6	3.7	91	21	AAG25492	Arabidopsis thalia
217	6	3.7	92	23	ABP39169	Staphylococcus epi
218	6	3.7	93	21	AAB32710	Eucalyptus grandis
219	6	3.7	99	22	ABB29270	Peptide #1921 enco
220	6	3.7	99	22	ABB34440	Peptide #1946 enco
221	6	3.7	99	22	ABB19850	Protein #1849 enco
222	6	3.7	99	22	AAM55226	Human brain expres
223	6	3.7	99	22	AAM67622	Human bone marrow
224	6	3.7	99	22	AAM15428	Peptide #1862 enco
225	6	3.7	99	22	AAM27914	Peptide #1951 enco
226	6	3.7	99	22	AAM03188	Peptide #1870 enco
227	6	3.7	99	23	ABG37203	Human peptide enco
228	6	3.7	101	22	AAB79316	Corynebacterium gl
229	6	3.7	103	23	AAE14722	Human lipocalin-2

230	6	3.7	106	22	AAU41451	Propionibacterium
231	6	3.7	107	22	AAB51265	Aspergillus niger
232	6	3.7	108	22	AAB49807	Murine endostatin
233	6	3.7	109	22	AAU56244	Propionibacterium
234	6	3.7	110	14	AAR33314	Variant IgE - muta
235	6	3.7	112	21	AAB56850	Human prostate can
236	6	3.7	112	21	AAY67502	PBR-associated pro
237	6	3.7	112	21	AAY54569	A starch binding d
238	6	3.7	114	23	ABP08643	Human ORFX protein
239	6	3.7	115	22	AAU66152	Propionibacterium
240	6	3.7	115	23	ABB55436	Lactococcus lactis
241	6	3.7	116	21	AAG42111	Zea mays protein f
242	6	3.7	116	22	ABG22364	Novel human diagno
243	6	3.7	116	22	ABG28656	Novel human diagno
244	6	3.7	117	21	AAG24365	Arabidopsis thalia
245	6	3.7	119	21	AAG26173	Zea mays protein f
246	6	3.7	119	21	AAG26748	Zea mays protein f
247	6	3.7	119	21	AAG35933	Zea mays protein f
248	6	3.7	120	20	AAY26348	Human secreted pro
249	6	3.7	120	21	AAY73399	Human secreted pro
250	6	3.7	120	22	AAM85593	Human immune/haema
251	6	3.7	120	23	ABB89655	Human polypeptide
252	6	3.7	121	22	AAU61958	Propionibacterium
253	6	3.7	121	22	AAO06409	Human polypeptide
254	6	3.7	124	18	AAW13537	Anti-melanoma anti
255	6	3.7	128	12	AAR12601	SIB 124 intestinal
256	6	3.7	128	22	ABB10756	Human ovarian and/
257	6	3.7	128	22	AAM94131	Human reproductive
258	6	3.7	128	23	AAM47605	Drosophila cell cy
259	6	3.7	131	21	AAG08948	Arabidopsis thalia
260	6	3.7	134	21	AAB40909	Human ORFX ORF673
261	6	3.7	134	22	AAU61435	Propionibacterium
262	6	3.7	134	22	ABG10015	Novel human diagno
263	6	3.7	134	23	ABP06756	Human ORFX protein
264	6	3.7	136	22	AAO13335	Human polypeptide
265	6	3.7	140	22	AAU44460	Propionibacterium
266	6	3.7	141	22	AAU40057	Propionibacterium
267	6	3.7	142	21	AAG42110	Zea mays protein f
268	6	3.7	144	22	ABG30154	Novel human diagno
269	6	3.7	144	22	AAY97767	I. scapularis Salp
270	6	3.7	145	21	AAG18093	Arabidopsis thalia
271	6	3.7	145	21	AAG02042	Human secreted pro
272	6	3.7	146	21	AAG08947	Arabidopsis thalia
273	6	3.7	146	22	ABG02557	Novel human diagno
274	6	3.7	146	23	ABP39485	Staphylococcus epi
275	6	3.7	147	23	ABB48575	Listeria monocytog
276	6	3.7	153	22	AAU66690	Propionibacterium
277	6	3.7	154	20	AAY13500	Tissue cement prot
278	6	3.7	154	22	AAM88089	Human immune/haema
279	6	3.7	154	23	AAM50381	Tick cement protei
280	6	3.7	157	23	AAM49151	Human visinin-like
281	6	3.7	163	22	AAO01930	Human polypeptide
282	6	3.7	165	22	AAG91560	C glutamicum prote
283	6	3.7	167	22	ABG26536	Novel human diagno
284	6	3.7	170	21	AAB25208	Eucalyptus grandis
285	6	3.7	171	21	AAG54995	Arabidopsis thalia
286	6	3.7	172	22	AAB79155	Corynebacterium gl
287	6	3.7	174	22	AAB80060	Corynebacterium gl
288	6	3.7	176	22	AAB78922	C. glutamicum SRT
289	6	3.7	176	22	AABB0155	Corynebacterium gl
290	6	3.7	176	23	ABB89620	Human polypeptide
291	6	3.7	182	18	AAW17521	Bacillus sp. heat
292	6	3.7	182	21	AAG08946	Arabidopsis thalia
293	6	3.7	182	22	AAM58208	Human brain expres
294	6	3.7	183	20	AAY38395	Human secreted pro
295	6	3.7	183	23	AAM49504	Mouse endostatin p
296	6	3.7	184	20	AAY18409	Endostatin protein
297	6	3.7	184	20	AAY08689	Murine endostatin
298	6	3.7	184	21	AAY70258	Murine angiogenesi
299	6	3.7	184	22	AAB49380	Murine endostatin
300	6	3.7	185	20	AAY06197	Anti-angiogenic en
301	6	3.7	185	22	ABG00746	Novel human diagno
302	6	3.7	188	22	AAU44663	Propionibacterium
303	6	3.7	191	21	AAB28398	Murine endostatin.
304	6	3.7	191	23	AAU77950	Amino acid sequenc
305	6	3.7	192	21	AAB58390	Lung cancer associ
306	6	3.7	192	22	AAG81744	S. epidermidis ope
307	6	3.7	193	21	AAG09652	Arabidopsis thalia
308	6	3.7	193	21	AAG50183	Arabidopsis thalia
309	6	3.7	199	22	AAU58769	Propionibacterium
310	6	3.7	200	21	AAG32357	Arabidopsis thalia
311	6	3.7	200	21	AAG48109	Arabidopsis thalia
312	6	3.7	201	22	AAB79315	Corynebacterium gl
313	6	3.7	201	22	AAB79716	Corynebacterium gl
314	6	3.7	202	21	AAY91299	Group B Streptococ
315	6	3.7	202	23	ABP29893	Streptococcus poly
316	6	3.7	203	21	AAG50144	Arabidopsis thalia
317	6	3.7	206	23	ABP28557	Streptococcus poly
318	6	3.7	207	22	AAE02031	Murine endostatin
319	6	3.7	207	22	AAB71930	Murine endostatin
320	6	3.7	212	21	AAG29412	Arabidopsis thalia
321	6	3.7	213	22	AAU38400	Salmonella typhi c
322	6	3.7	213	22	AAU56474	Propionibacterium
323	6	3.7	214	22	AAU43301	Propionibacterium
324	6	3.7	217	21	AAG29411	Arabidopsis thalia
325	6	3.7	218	20	AAY59857	Human normal uteru
326	6	3.7	218	20	AAY08691	Murine gene therap
327	6	3.7	219	21	AAG29410	Arabidopsis thalia
328	6	3.7	221	22	AAU60655	Propionibacterium
329	6	3.7	221	23	ABB55237	Lactococcus lactis
330	6	3.7	224	23	ABP26451	Streptococcus poly
331	6	3.7	226	20	AAY37114	Chlamydia trachoma
332	6	3.7	226	23	ABP27127	Streptococcus poly
333	6	3.7	226	23	ABB54817	Lactococcus lactis
334	6	3.7	228	22	ABG15076	Novel human diagno
335	6	3.7	228	23	ABP26927	Streptococcus poly
336	6	3.7	230	22	ABB60542	Drosophila melanog
337	6	3.7	230	22	AAG83110	S. epidermidis ope
338	6	3.7	232	21	AAG32356	Arabidopsis thalia
339	6	3.7	232	21	AAG48108	Arabidopsis thalia
340	6	3.7	233	21	AAG32937	Arabidopsis thalia
341	6	3.7	234	22	ABB63157	Drosophila melanog
342	6	3.7	235	21	AAG39986	Arabidopsis thalia
343	6	3.7	235	22	AAB96157	Putative P. abyssi
344	6	3.7	237	22	AAB82582	Human prothrombina
345	6	3.7	239	22	ABG29657	Novel human diagno
346	6	3.7	240	22	AAB82592	Human prothrombina
347	6	3.7	243	14	AAR34705	JSC1 precursor. R
348	6	3.7	243	19	AAW73161	Rat amphiregulin.
349	6	3.7	245	22	AAB67617	Human leukocyte an
350	6	3.7	245	22	AAB67618	Human leukocyte an
351	6	3.7	245	22	AAB67619	Human leukocyte an
352	6	3.7	245	22	AAB67620	Human leukocyte an
353	6	3.7	245	22	AAB67621	Human leukocyte an
354	6	3.7	245	22	AAB67622	Human leukocyte an
355	6	3.7	245	22	AAB67623	Human leukocyte an
356	6	3.7	247	18	AAW08475	Porcine trypsinoge
357	6	3.7	247	22	AAG82186	S. epidermidis ope
358	6	3.7	248	21	AAB56222	Human secreted pro
359	6	3.7	248	21	AAG48866	Arabidopsis thalia
360	6	3.7	249	21	AAY52296	Human isomerase ho
361	6	3.7	249	23	ABP44395	Human BLyS binding
362	6	3.7	249	23	ABP44407	Human BLyS binding
363	6	3.7	251	21	AAG32936	Arabidopsis thalia
364	6	3.7	251	22	ABG18585	Novel human diagno
365	6	3.7	251	22	ABG21428	Novel human diagno
366	6	3.7	251	22	ABG25269	Novel human diagno
367	6	3.7	252	22	AAE11019	Recombinant viral
368	6	3.7	252	23	ABB08111	FIV-Oma3 vif prote
369	6	3.7	253	22	AAB73499	Human transferase
370	6	3.7	260	21	AAG39880	Arabidopsis thalia
371	6	3.7	261	22	AAO05937	Human polypeptide
372	6	3.7	262	15	AAR60478	Clavulanic acid re
373	6	3.7	262	23	ABB53604	Lactococcus lactis
374	6	3.7	269	22	AAG82032	S. epidermidis ope
375	6	3.7	270	22	AAE01961	Arabidopsis thalia

376	6	3.7	270	22	AAB67838	Amino acid sequenc
377	6	3.7	271	22	AAU87132	Novel central nerv
378	6	3.7	271	23	AAG66039	scFv 1b4 antibody
379	6	3.7	272	21	AAB24140	Cucumber antimicro
380	6	3.7	272	22	AAG91844	C glutamicum prote
381	6	3.7	273	21	AAG39879	Arabidopsis thalia
382	6	3.7	273	22	ABB70029	Drosophila melanog
383	6	3.7	273	23	ABB92422	Herbicidally activ
384	6	3.7	275	21	AAG25357	Arabidopsis thalia
385	6	3.7	275	23	ABB32472	Staphylococcus aur
386	6	3.7	280	21	AAG32935	Arabidopsis thalia
387	6	3.7	281	18	AAW27560	Consensus single c
388	6	3.7	281	21	AAG39878	Arabidopsis thalia
389	6	3.7	281	22	ABB11315	Human HSPP-124 pro
390	6	3.7	284	23	ABB32484	Staphylococcus aur
391	6	3.7	291	21	AAG39985	Arabidopsis thalia
392	6	3.7	293	21	AAB29468	Cucumber xylem sap
393	6	3.7	295	22	ABG06263	Novel human diagno
394	6	3.7	297	15	AAR48736	G-protein coupled
395	6	3.7	297	15	AAR48737	G-protein coupled
396	6	3.7	297	17	AAW02708	G-protein coupled
397	6	3.7	297	17	AAW02709	G-protein coupled
398	6	3.7	297	21	AAG30795	Arabidopsis thalia
399	6	3.7	297	23	ABP26875	Streptococcus poly
400	6	3.7	297	23	ABB90824	Herbicidally activ
401	6	3.7	301	22	ABB63512	Drosophila melanog
402	6	3.7	302	21	AAG39230	Arabidopsis thalia
403	6	3.7	302	22	AAU40372	Propionibacterium
404	6	3.7	302	22	AAE02526	Arabidopsis thalia
405	6	3.7	305	21	AAG05315	Arabidopsis thalia
406	6	3.7	305	21	AAG39980	Arabidopsis thalia
407	6	3.7	305	22	AAU58771	Propionibacterium
408	6	3.7	306	22	AAB82593	Human prothrombina
409	6	3.7	306	23	ABP51325	Human MDDT SEQ ID
410	6	3.7	307	23	ABP38393	Staphylococcus epi
411	6	3.7	308	21	AAG43295	Arabidopsis thalia
412	6	3.7	309	20	AAY37181	Protein involved i
413	6	3.7	314	22	ABG10894	Novel human diagno
414	6	3.7	315	22	ABG11079	Novel human diagno
415	6	3.7	315	22	AAE10324	Human transporter
416	6	3.7	315	22	AAB94269	Human protein sequ
417	6	3.7	315	22	AAB94664	Human protein sequ
418	6	3.7	318	12	AAR15480	NF-YA. Homo sapie
419	6	3.7	318	21	AAG09036	Arabidopsis thalia
420	6	3.7	322	22	AAG92292	C glutamicum prote
421	6	3.7	322	22	AAG93157	C glutamicum prote
422	6	3.7	323	18	AAW34200	Streptomyces efflu
423	6	3.7	323	19	AAW55801	Streptomyces roseo
424	6	3.7	326	21	AAG05752	Arabidopsis thalia
425	6	3.7	326	22	AAU14275	Human novel protei
426	6	3.7	326	22	AAU14511	Human novel protei
427	6	3.7	326	22	AAB82590	Human prothrombina
428	6	3.7	331	22	AAU50916	Propionibacterium
429	6	3.7	331	22	AAG89937	C glutamicum prote
430	6	3.7	335	21	AAB56997	Human prostate can
431	6	3.7	335	22	ABB66970	Drosophila melanog
432	6	3.7	335	22	AAG90139	C glutamicum prote
433	6	3.7	344	23	AAE24030	Human HPR2 spliced
434	6	3.7	345	22	ABB63738	Drosophila melanog
435	6	3.7	346	22	AAG65608	Human oocyte speci
436	6	3.7	347	22	AAU41440	Propionibacterium
437	6	3.7	347	23	ABB04964	Human adenovirus t
438	6	3.7	348	22	AAB96784	Putative P. abyssi
439	6	3.7	350	21	AAG48107	Arabidopsis thalia
440	6	3.7	350	22	ABB59032	Drosophila melanog
441	6	3.7	352	21	AAG32355	Arabidopsis thalia
442	6	3.7	352	22	AAE08011	Rat chimeric rNPY5
443	6	3.7	353	23	AAU79045	Human novel human
444	6	3.7	354	22	AAU01214	Enterococcus faeci
445	6	3.7	354	23	AAU11303	Thermotoga maritim
446	6	3.7	355	22	ABG03603	Novel human diagno
447	6	3.7	356	21	AAG05751	Arabidopsis thalia
448	6	3.7	358	21	AAB16412	Pinus radiata diph
449	6	3.7	358	21	AAG25356	Arabidopsis thalia
450	6	3.7	360	21	AAG09035	Arabidopsis thalia
451	6	3.7	360	21	AAG19661	Arabidopsis thalia
452	6	3.7	361	21	AAG09034	Arabidopsis thalia
453	6	3.7	361	21	AAG19660	Arabidopsis thalia
454	6	3.7	362	22	AAE01937	Arabidopsis thalia
455	6	3.7	365	21	AAB56813	Human prostate can
456	6	3.7	365	21	AAG06592	Arabidopsis thalia
457	6	3.7	365	21	AAG45295	Arabidopsis thalia
458	6	3.7	365	23	AAU11305	Thermotoga maritim
459	6	3.7	366	23	ABP39520	Staphylococcus epi
460	6	3.7	366	23	AAU10816	Human Type II GnRH
461	6	3.7	367	21	AAG39984	Arabidopsis thalia
462	6	3.7	367	23	ABB92915	Herbicidally activ
463	6	3.7	368	22	AAB82589	Human prothrombina
464	6	3.7	368	23	AAU11307	Variant fusion pro
465	6	3.7	369	23	AAU11306	Fusion protein fro
466	6	3.7	370	22	AAG70708	S cerevisiae apopt
467	6	3.7	371	22	AAG83100	S. epidermidis ope
468	6	3.7	372	21	AAG46626	Arabidopsis thalia
469	6	3.7	372	22	AAE07865	Novel human protei
470	6	3.7	376	21	AAG06591	Arabidopsis thalia
471	6	3.7	376	21	AAG45294	Arabidopsis thalia
472	6	3.7	376	23	ABP39414	Staphylococcus epi
473	6	3.7	376	23	AAU10815	Human Type II GnRH
474	6	3.7	377	22	ABB61905	Drosophila melanog
475	6	3.7	377	23	AAU10814	Human Type II GnRH
476	6	3.7	377	23	AAU10819	Human Type II GnRH
477	6	3.7	378	22	AAM39027	Human polypeptide
478	6	3.7	379	20	AAY25768	Human secreted pro
479	6	3.7	379	21	AAB56086	Human secreted pro
480	6	3.7	379	23	AAU10813	Human Type II GnRH
481	6	3.7	380	21	AAG25355	Arabidopsis thalia
482	6	3.7	380	23	AAU10812	Marmoset Type II G
483	6	3.7	381	21	AAG14473	Arabidopsis thalia
484	6	3.7	382	22	ABB76934	Rat VG41. Rattus
485	6	3.7	383	21	AAG05750	Arabidopsis thalia
486	6	3.7	383	23	AAU11304	Fusion protein fro
487	6	3.7	385	21	AAG45293	Arabidopsis thalia
488	6	3.7	386	20	AAY38873	Neisseria meningit
489	6	3.7	386	21	AAG06590	Arabidopsis thalia
490	6	3.7	386	23	ABB53527	Lactococcus lactis
491	6	3.7	387	22	AAB95516	Human protein sequ
492	6	3.7	388	21	AAG08797	Arabidopsis thalia
493	6	3.7	388	21	AAG12961	Arabidopsis thalia
494	6	3.7	388	22	ABG07515	Novel human diagno
495	6	3.7	388	22	ABG20786	Novel human diagno
496	6	3.7	388	22	AAB82584	Human prothrombina
497	6	3.7	388	22	AAY72622	Human angiopoietin
498	6	3.7	388	23	ABB77546	Human polypeptide
499	6	3.7	389	21	AAY74909	Neisseria meningit
500	6	3.7	391	21	AAG46625	Arabidopsis thalia
501	6	3.7	391	21	AAY70005	Casein kinase II a
502	6	3.7	394	21	AAG05314	Arabidopsis thalia
503	6	3.7	394	21	AAG39979	Arabidopsis thalia
504	6	3.7	394	23	ABB90861	Herbicidally activ
505	6	3.7	395	22	AAE08013	Rat chimeric rNPY5
506	6	3.7	398	21	AAG20852	Arabidopsis thalia
507	6	3.7	398	21	AAG43104	Arabidopsis thalia
508	6	3.7	398	22	ABB59789	Drosophila melanog
509	6	3.7	401	21	AAG12960	Arabidopsis thalia
510	6	3.7	402	21	AAG39978	Arabidopsis thalia
511	6	3.7	403	21	AAG05313	Arabidopsis thalia
512	6	3.7	403	21	AAG14472	Arabidopsis thalia
513	6	3.7	404	23	ABP27853	Streptococcus poly
514	6	3.7	405	20	AAY38875	Neisseria meningit
515	6	3.7	405	20	AAY38876	Neisseria gonorrho
516	6	3.7	405	22	AAG92658	C glutamicum prote
517	6	3.7	405	22	AAB79146	Corynebacterium gl
518	6	3.7	406	22	AAB93320	Human protein sequ
519	6	3.7	406	22	AAB68523	Human GTP-binding
520	6	3.7	407	20	AAY38874	Neisseria meningit
521	6	3.7	407	21	AAG08796	Arabidopsis thalia

522	6	3.7	410	22	ABB65286	Drosophila melanog
523	6	3.7	417	13	AAR21573	HCV CKS-NS3 - pHCV
524	6	3.7	417	14	AAR33640	HCV CKS-NS3 recomb
525	6	3.7	417	14	AAR33582	HCV CKS-NS3 recomb
526	6	3.7	417	14	AAR33602	HCV CKS-NS3 recomb
527	6	3.7	417	22	AAB69001	HCV recombinant an
528	6	3.7	418	21	AAG08795	Arabidopsis thalia
529	6	3.7	418	21	AAG43294	Arabidopsis thalia
530	6	3.7	421	22	AAU23239	Novel human enzyme
531	6	3.7	423	21	AAG14471	Arabidopsis thalia
532	6	3.7	423	21	AAG43293	Arabidopsis thalia
533	6	3.7	425	21	AAG57212	Arabidopsis thalia
534	6	3.7	429	20	AAY07042	Breast cancer asso
535	6	3.7	430	20	AAY35283	Chlamydia pneumoni
536	6	3.7	431	22	AAG82736	S. epidermidis ope
537	6	3.7	431	23	AAM50319	Proton-dependent o
538	6	3.7	433	21	AAG57211	Arabidopsis thalia
539	6	3.7	435	21	AAG46609	Arabidopsis thalia
540	6	3.7	435	23	ABB54160	Lactococcus lactis
541	6	3.7	436	22	ABG24048	Novel human diagno
542	6	3.7	445	18	AAW27602	Rat neuropeptide Y
543	6	3.7	445	18	AAW27603	Rat neuropeptide Y
544	6	3.7	445	18	AAW15232	Rat neuropeptide Y
545	6	3.7	445	20	AAY14552	Rat neuropeptide Y
546	6	3.7	445	20	AAY14553	Rat neuropeptide Y
547	6	3.7	445	21	AAY52577	Chimeric rat/human
548	6	3.7	445	21	AAY52579	Rat NPY (neuropept
549	6	3.7	446	21	AAG46709	Arabidopsis thalia
550	6	3.7	446	22	ABB11726	Human semaphorin Y
551	6	3.7	446	22	AAE07864	Novel human protei
552	6	3.7	447	22	AAY72393	Rice LHT1 encoded
553	6	3.7	451	23	ABP27531	Streptococcus poly
554	6	3.7	453	19	AAW62857	Agrocybe pediades
555	6	3.7	453	20	AAY39907	A. pediades phytas
556	6	3.7	454	21	AAG46608	Arabidopsis thalia
557	6	3.7	455	20	AAY38617	Neisseria gonorrho
558	6	3.7	455	21	AAY74908	Neisseria gonorrhe
559	6	3.7	455	21	AAY74910	Neisseria meningit
560	6	3.7	456	17	AAR95940	Rat Y5 receptor.
561	6	3.7	456	18	AAW29446	Rat hypothalamic n
562	6	3.7	456	18	AAW29412	Rat hypothalamic n
563	6	3.7	456	19	AAW37092	Rattus norvegicus
564	6	3.7	456	20	AAY57460	Rat hypothalamic Y
565	6	3.7	459	22	ABB61416	Drosophila melanog
566	6	3.7	460	23	AAU91167	Pantothenate kinas
567	6	3.7	461	21	AAG20851	Arabidopsis thalia
568	6	3.7	461	21	AAG43103	Arabidopsis thalia
569	6	3.7	466	18	AAW15233	Mouse neuropeptide
570	6	3.7	466	21	AAY53635	A bone marrow secr
571	6	3.7	466	23	ABP38510	Staphylococcus epi
572	6	3.7	467	21	AAG20850	Arabidopsis thalia
573	6	3.7	467	21	AAG43102	Arabidopsis thalia
574	6	3.7	472	22	ABB60324	Drosophila melanog
575	6	3.7	473	22	AAG92443	C glutamicum prote
576	6	3.7	475	22	ABB76935	Rat VG51. Rattus
577	6	3.7	478	20	AAY14051	G. oxydans D-sorbi
578	6	3.7	478	21	AAB42226	Human ORFX ORF1990
579	6	3.7	478	21	AAY57603	Human protein SEQ
580	6	3.7	482	22	ABB66681	Drosophila melanog
581	6	3.7	483	22	ABB66467	Drosophila melanog
582	6	3.7	484	21	AAG47652	Arabidopsis thalia
583	6	3.7	485	23	ABB47906	Listeria monocytog
584	6	3.7	486	22	ABB68106	Drosophila melanog
585	6	3.7	487	21	AAB26490	Soybean aminopepti
586	6	3.7	487	21	AAG47632	Arabidopsis thalia
587	6	3.7	488	22	AAG89962	C glutamicum prote
588	6	3.7	488	22	AAB80058	Corynebacterium gl
589	6	3.7	492	23	ABB07836	Human protein iden
590	6	3.7	494	21	AAG05024	Arabidopsis thalia
591	6	3.7	496	22	ABB52510	Escherichia coli p
592	6	3.7	499	22	AAE08012	Rat chimeric rNPY5
593	6	3.7	508	22	AAE08010	Mouse chimeric mNP
594	6	3.7	509	19	AAW40480	Rat SH2 binding pr
595	6	3.7	512	21	AAG48894	Arabidopsis thalia
596	6	3.7	514	22	ABB52689	Escherichia coli p
597	6	3.7	514	22	AAG81103	Mycobacterium tube
598	6	3.7	515	22	AAU02903	Angiotensin conver
599	6	3.7	516	18	AAW17974	Coprinus cinereus
600	6	3.7	516	22	ABB61121	Drosophila melanog
601	6	3.7	521	20	AAY39780	Interleukin-18 rec
602	6	3.7	522	21	AAG46708	Arabidopsis thalia
603	6	3.7	523	23	ABB07814	Chicken heparanase
604	6	3.7	524	21	AAG47651	Arabidopsis thalia
605	6	3.7	527	21	AAG47631	Arabidopsis thalia
606	6	3.7	528	22	ABG13658	Novel human diagno
607	6	3.7	529	22	ABB69774	Drosophila melanog
608	6	3.7	530	20	AAY31650	Brevibacterium fla
609	6	3.7	530	20	AAY31649	Brevibacterium fla
610	6	3.7	530	20	AAY22646	Wild type 3-PGDH p
611	6	3.7	530	20	AAY22647	Mutant 3-PGDH prot
612	6	3.7	530	22	AAG91161	C glutamicum prote
613	6	3.7	531	21	AAG11571	Arabidopsis thalia
614	6	3.7	533	22	AAG90402	C glutamicum prote
615	6	3.7	533	23	ABB54316	Lactococcus lactis
616	6	3.7	534	21	AAG05023	Arabidopsis thalia
617	6	3.7	538	22	AAM79344	Human protein SEQ
618	6	3.7	540	19	AAW77145	Human interleukin-
619	6	3.7	540	20	AAY39786	Interleukin-18 rec
620	6	3.7	541	18	AAW31907	Human receptor pro
621	6	3.7	541	20	AAY06613	Human interleukin-
622	6	3.7	541	23	AAE14731	Human IL-18 recept
623	6	3.7	541	23	AAE16583	Human interleukin-
624	6	3.7	544	22	AAB48054	Signal transductio
625	6	3.7	552	22	AAU43492	Propionibacterium
626	6	3.7	564	22	ABG06265	Novel human diagno
627	6	3.7	566	22	AAU40040	Propionibacterium
628	6	3.7	570	17	AAR96300	FokI insertion mut
629	6	3.7	570	22	AAG91849	C glutamicum prote
630	6	3.7	573	17	AAR96299	FokI insertion mut
631	6	3.7	573	22	ABB70865	Drosophila melanog
632	6	3.7	576	16	AAR71374	Human GRK #2. Hom
633	6	3.7	576	22	AAB84262	Amino acid sequenc
634	6	3.7	576	23	AAU79932	Human G protein-co
635	6	3.7	577	15	AAR58759	FokI amino acid se
636	6	3.7	579	16	AAR72810	FokI. Flavobacter
637	6	3.7	580	20	AAY08692	Murine gene therap
638	6	3.7	581	17	AAR88594	FokI insertion mut
639	6	3.7	581	17	AAR88593	FokI insertion mut
640	6	3.7	582	10	AAP93666	Hemagglutinin-neur
641	6	3.7	582	22	AAB67478	Amino acid sequenc
642	6	3.7	585	17	AAR88595	FokI insertion mut
643	6	3.7	586	22	ABB66107	Drosophila melanog
644	6	3.7	589	17	AAR88596	FokI insertion mut
645	6	3.7	592	20	AAY38618	Neisseria gonorrho
646	6	3.7	592	20	AAY38615	Neisseria meningit
647	6	3.7	592	20	AAY38616	Neisseria meningit
648	6	3.7	592	21	AAY74911	Neisseria gonorrhe
649	6	3.7	592	21	AAY74912	Neisseria meningit
650	6	3.7	592	21	AAY74913	Neisseria meningit
651	6	3.7	592	23	AAU91166	Pantothenate kinas
652	6	3.7	592	23	AAU91169	Pantothenate kinas
653	6	3.7	592	23	ABB54353	Lactococcus lactis
654	6	3.7	595	17	AAR88599	FokI insertion mut
655	6	3.7	595	22	AAU87547	Novel central nerv
656	6	3.7	595	22	AAU23294	Novel human enzyme
657	6	3.7	596	17	AAR88597	FokI insertion mut
658	6	3.7	596	19	AAW77099	P. chrysogenum bet
659	6	3.7	600	17	AAR88598	FokI insertion mut
660	6	3.7	600	19	AAW39080	S. lepidophylla tr
661	6	3.7	603	18	AAW17889	Photorhabdus lumin
662	6	3.7	603	19	AAW56555	Toxin TcaAiii, enc
663	6	3.7	605	22	AAM93959	Human polypeptide,
664	6	3.7	605	23	AAU76819	Corn contig plasti
665	6	3.7	613	22	ABB71965	Drosophila melanog
666	6	3.7	615	22	AAM40056	Human polypeptide
667	6	3.7	616	19	AAW55976	Aspergillus awamor

668	6	3.7	616	19	AAW55977	Aspergillus awamor
669	6	3.7	616	19	AAW55979	Aspergillus awamor
670	6	3.7	616	19	AAW55980	Aspergillus awamor
671	6	3.7	616	21	AAB15176	Aspergillus awamor
672	6	3.7	616	21	AAB15178	Aspergillus awamor
673	6	3.7	616	21	AAB15179	A.awamori mutant g
674	6	3.7	616	21	AAB15180	Aspergillus awamor
675	6	3.7	616	21	AAB15181	Aspergillus awamor
676	6	3.7	616	21	AAB15184	A.awamori mutant g
677	6	3.7	616	21	AAG46707	Arabidopsis thalia
678	6	3.7	621	19	AAW55978	Aspergillus awamor
679	6	3.7	621	21	AAB15182	A.awamori mutant g
680	6	3.7	621	21	AAB15183	A.awamori mutant g
681	6	3.7	621	23	AAU11270	Human coagulation
682	6	3.7	621	23	AAU11271	Human coagulation
683	6	3.7	630	9	AAP81876	Sequence of amylog
684	6	3.7	631	22	ABG22173	Novel human diagno
685	6	3.7	632	16	AAR71373	Human GRK #1. Hom
686	6	3.7	637	15	AAR63673	Glutamylcysteine-s
687	6	3.7	639	5	AAP40212	Sequence encoded b
688	6	3.7	640	21	AAY77741	A. niger G1 glucoa
689	6	3.7	640	22	AAB61905	A. niger protein s
690	6	3.7	650	22	AAE07867	Novel human protei
691	6	3.7	655	23	AAU76822	Rice clone rdilc.p
692	6	3.7	658	21	AAG31406	Arabidopsis thalia
693	6	3.7	663	22	ABG08738	Novel human diagno
694	6	3.7	669	22	ABB61302	Drosophila melanog
695	6	3.7	669	22	AAB47564	Protease PRTS-6.
696	6	3.7	669	23	AAU79044	Human novel human
697	6	3.7	669	23	AAU79046	Human novel human
698	6	3.7	669	23	AAU72883	Human aspartyl pro
699	6	3.7	675	18	AAW06449	Brassica napus FCA
700	6	3.7	679	21	AAB58294	Lung cancer associ
701	6	3.7	684	20	AAY25114	Mouse alpha1 (XVII
702	6	3.7	687	21	AAG31405	Arabidopsis thalia
703	6	3.7	691	21	AAG51807	Arabidopsis thalia
704	6	3.7	692	22	AAB65619	Novel protein kina
705	6	3.7	695	22	ABB59756	Drosophila melanog
706	6	3.7	701	21	AAG31685	Arabidopsis thalia
707	6	3.7	701	22	AAG92442	C glutamicum prote
708	6	3.7	720	21	AAG51806	Arabidopsis thalia
709	6	3.7	724	22	AAE07866	Novel human protei
710	6	3.7	727	21	AAG31404	Arabidopsis thalia
711	6	3.7	727	23	ABB93080	Herbicidally activ
712	6	3.7	730	21	AAG31684	Arabidopsis thalia
713	6	3.7	734	22	ABB63238	Drosophila melanog
714	6	3.7	734	22	ABB63239	Drosophila melanog
715	6	3.7	747	22	ABB67181	Drosophila melanog
716	6	3.7	751	23	AAE14444	Human drug metabol
717	6	3.7	759	22	ABG07769	Novel human diagno
718	6	3.7	759	22	ABG08475	Novel human diagno
719	6	3.7	760	21	AAG51805	Arabidopsis thalia
720	6	3.7	766	23	AAE21719	Human PKIN-14 prot
721	6	3.7	766	23	AAG66161	Human serine/threo
722	6	3.7	766	23	ABB92836	Herbicidally activ
723	6	3.7	770	20	AAY13456	Amino acid sequenc
724	6	3.7	770	21	AAG31683	Arabidopsis thalia
725	6	3.7	771	22	ABB65630	Drosophila melanog
726	6	3.7	771	22	AAE07869	Novel human protei
727	6	3.7	785	22	AAG90090	C glutamicum prote
728	6	3.7	791	22	AAM39420	Human polypeptide
729	6	3.7	793	22	ABB71688	Drosophila melanog
730	6	3.7	796	21	AAG30685	Arabidopsis thalia
731	6	3.7	802	22	AAB47592	Human DHDR-1. Hom
732	6	3.7	803	22	AAU87246	Novel central nerv
733	6	3.7	806	22	ABG21109	Novel human diagno
734	6	3.7	808	20	AAY49873	Arabidopsis ionotr
735	6	3.7	808	23	ABB92200	Herbicidally activ
736	6	3.7	809	22	ABB59391	Drosophila melanog
737	6	3.7	809	23	ABB91284	Herbicidally activ
738	6	3.7	827	18	AAW07686	Ionotropic glutama
739	6	3.7	831	22	AAG91714	C glutamicum prote
740	6	3.7	837	22	AAM41206	Human polypeptide
741	6	3.7	838	23	ABB08523	Amino acid sequenc
742	6	3.7	839	13	AAR22275	Bovine liver GPI-P
743	6	3.7	839	16	AAR75109	Glycosyl-phosphati
744	6	3.7	845	22	AAE07868	Novel human protei
745	6	3.7	873	22	AAG82971	S. epidermidis ope
746	6	3.7	895	22	ABB63617	Drosophila melanog
747	6	3.7	897	23	ABP38755	Staphylococcus epi
748	6	3.7	900	23	ABB93716	Herbicidally activ
749	6	3.7	921	15	AAR56487	TATA-binding prote
750	6	3.7	921	17	AAW06077	Drosophila TATA-bi
751	6	3.7	921	18	AAW25028	TATA-binding prote
752	6	3.7	921	22	ABB61528	Drosophila melanog
753	6	3.7	921	22	ABB66055	Drosophila melanog
754	6	3.7	924	22	ABG30172	Novel human diagno
755	6	3.7	924	22	AAB93084	Human protein sequ
756	6	3.7	924	22	AAB95845	Human protein sequ
757	6	3.7	925	21	AAG30684	Arabidopsis thalia
758	6	3.7	925	23	ABB93251	Herbicidally activ
759	6	3.7	930	19	AAW57260	Human semaphorin Y
760	6	3.7	930	20	AAY35052	Chlamydia pneumoni
761	6	3.7	930	20	AAW88424	Chlamydia pneumoni
762	6	3.7	930	21	AAY90240	Chlamydia antigen
763	6	3.7	930	23	ABB90548	Chlamydia pneumoni
764	6	3.7	936	21	AAG30683	Arabidopsis thalia
765	6	3.7	950	21	AAY53899	Amino acid sequenc
766	6	3.7	958	22	ABG06858	Novel human diagno
767	6	3.7	968	22	ABB68484	Drosophila melanog
768	6	3.7	973	18	AAW41815	A. thaliana trehal
769	6	3.7	973	22	AAM41045	Human polypeptide
770	6	3.7	975	22	ABG24301	Novel human diagno
771	6	3.7	987	22	ABB63604	Drosophila melanog
772	6	3.7	994	19	AAW44844	S. lepidophylla tr
773	6	3.7	994	22	ABG25520	Novel human diagno
774	6	3.7	997	22	ABB59064	Drosophila melanog
775	6	3.7	1027	18	AAW17057	Candida albicans c
776	6	3.7	1054	22	ABB57993	Drosophila melanog
777	6	3.7	1057	23	ABB92982	Herbicidally activ
778	6	3.7	1058	22	ABB71401	Drosophila melanog
779	6	3.7	1081	22	ABB66277	Drosophila melanog
780	6	3.7	1081	22	ABB71824	Drosophila melanog
781	6	3.7	1095	18	AAW17888	Photorhabdus lumin
782	6	3.7	1095	19	AAW56547	Toxin TcaA, encode
783	6	3.7	1116	22	ABG22811	Novel human diagno
784	6	3.7	1134	23	ABB93262	Herbicidally activ
785	6	3.7	1164	22	ABB61495	Drosophila melanog
786	6	3.7	1166	20	AAY08643	S. aureus SdrE pro
787	6	3.7	1174	22	ABB65304	Drosophila melanog
788	6	3.7	1219	22	ABB65388	Drosophila melanog
789	6	3.7	1225	19	AAW52289	Homo sapiens cdo t
790	6	3.7	1226	21	AAG42239	Arabidopsis thalia
791	6	3.7	1278	22	AAM39259	Human polypeptide
792	6	3.7	1288	18	AAW26328	Mouse alpha-1 coll
793	6	3.7	1288	20	AAW92297	Mouse alpha-1 (XVI
794	6	3.7	1297	22	ABB64155	Drosophila melanog
795	6	3.7	1299	22	ABB70025	Drosophila melanog
796	6	3.7	1310	22	ABB60913	Drosophila melanog
797	6	3.7	1321	21	AAG42238	Arabidopsis thalia
798	6	3.7	1351	21	AAG42237	Arabidopsis thalia
799	6	3.7	1352	21	AAG31776	Arabidopsis thalia
800	6	3.7	1360	22	AAU11802	Human protein kina
801	6	3.7	1374	21	AAB11934	Human MEKK5. Homo
802	6	3.7	1375	18	AAW27283	Apoptosis inducing
803	6	3.7	1377	23	ABB77532	Ring finger protei
804	6	3.7	1400	20	AAW82791	Human RON receptor
805	6	3.7	1404	23	AAE20581	Proliferation pote
806	6	3.7	1406	21	AAG31775	Arabidopsis thalia
807	6	3.7	1407	21	AAG31774	Arabidopsis thalia
808	6	3.7	1407	22	AAB86123	A. thaliana formyl
809	6	3.7	1446	22	ABB63929	Drosophila melanog
810	6	3.7	1462	23	ABB97521	Novel human protei
811	6	3.7	1499	22	ABB58409	Drosophila melanog
812	6	3.7	1508	22	AAB50676	C. elegans alpha-2
813	6	3.7	1519	22	AAB50677	C. elegans alpha-2

814	6	3.7	1528	17	AAR95333	Manduca sexta Baci
815	6	3.7	1528	20	AAW90182	Manduca sexta BT t
816	6	3.7	1566	22	ABB64864	Drosophila melanog
817	6	3.7	1617	22	AAE07870	Novel human protei
818	6	3.7	1635	23	ABB90958	Herbicidally activ
819	6	3.7	1653	22	AAG98354	Escherichia coli p
820	6	3.7	1691	22	AAE10609	Human novel KIAA12
821	6	3.7	1691	22	AAE07863	Novel human protei
822	6	3.7	1707	22	AAG78760	Hepatitis E virus
823	6	3.7	1752	23	ABB05030	Bifidobacterium bi
824	6	3.7	1897	21	AAB19712	Human protein tyro
825	6	3.7	1897	21	AAY81785	Human protein tyro
826	6	3.7	1897	21	AAY56100	LAR tyrosine phosp
827	6	3.7	1907	22	AAU14143	Human novel protei
828	6	3.7	1916	22	ABB62423	Drosophila melanog
829	6	3.7	1916	22	ABB66089	Drosophila melanog
830	6	3.7	1940	22	ABB64070	Drosophila melanog
831	6	3.7	1963	22	ABB58053	Drosophila melanog
832	6	3.7	2070	22	ABG03703	Novel human diagno
833	6	3.7	2088	22	AAM80124	Human protein SEQ
834	6	3.7	2090	22	ABB64682	Drosophila melanog
835	6	3.7	2114	22	ABG09076	Novel human diagno
836	6	3.7	2166	22	AAM79140	Human protein SEQ
837	6	3.7	2209	23	AAU85609	Lung tumour protei
838	6	3.7	2215	22	AAG64030	Human polypeptide
839	6	3.7	2356	22	ABG09072	Novel human diagno
840	6	3.7	2408	13	AAR24306	Translation of ORF
841	6	3.7	2496	18	AAW36093	Mutant YLR087c pro
842	6	3.7	2713	18	AAW19654	ATM mutant C8140T.
843	6	3.7	2759	18	AAW19652	ATM mutant 8269del
844	6	3.7	2766	18	AAW19651	ATM mutant 8283del
845	6	3.7	2768	18	AAW19650	ATM mutant G8307A.
846	6	3.7	2932	18	AAW19675	ATM mutant 2467del
847	6	3.7	2942	18	AAW22050	Saccharomyces cere
848	6	3.7	2987	18	AAW19649	ATM mutant 8946ins
849	6	3.7	2989	18	AAW19688	ATM mutant 1407del
850	6	3.7	2998	18	AAW19691	ATM mutant 3403del
851	6	3.7	2998	18	AAW19673	ATM mutant 3403del
852	6	3.7	3001	18	AAW19692	ATM mutant 4612del
853	6	3.7	3001	18	AAW19670	ATM mutant 4612del
854	6	3.7	3003	18	AAW19658	ATM mutant 7630del
855	6	3.7	3005	18	AAW19653	ATM mutant 8269del
856	6	3.7	3005	18	AAW19648	ATM mutant 9001del
857	6	3.7	3021	18	AAW19661	ATM mutant 6348del
858	6	3.7	3026	18	AAW19676	ATM mutant 2377del
859	6	3.7	3046	18	AAW19690	ATM mutant C9140T.
860	6	3.7	3053	18	AAW19694	ATM mutant 7636del
861	6	3.7	3054	18	AAW19702	ATM mutant 7278del
862	6	3.7	3054	18	AAW19695	ATM mutant 7279del
863	6	3.7	3055	18	AAW19697	ATM mutant 5435del
864	6	3.7	3055	18	AAW19693	ATM mutant 8578del
865	6	3.7	3056	18	AAW06234	ATM protein. Homo
866	6	3.7	3056	18	AAW07655	Ataxia-telangiecta
867	6	3.7	3056	18	AAW19733	Cell cycle checkpo
868	6	3.7	3056	18	AAW19698	ATM mutant E2904G.
869	6	3.7	3056	18	AAW19699	ATM mutant Leu43Pr
870	6	3.7	3056	19	AAW37133	Ataxia-telangiecta
871	6	3.7	3056	20	AAW84269	Human ataxia telan
872	6	3.7	3056	22	AAB62280	ATM open reading f
873	6	3.7	3056	22	AAB68895	Human ATM protein.
874	6	3.7	3056	22	AAG67443	Amino acid sequenc
875	6	3.7	3056	22	AAG67622	Amino acid sequenc
876	6	3.7	3057	22	ABG05614	Novel human diagno
877	6	3.7	3057	22	AAB68896	Human ATM protein.
878	6	3.7	3059	18	AAW19696	ATM mutant 5319ins
879	6	3.7	3066	18	AAW36178	Murine Ataxia-tela
880	6	3.7	3085	18	AAW19701	ATM mutant G9170C.
881	6	3.7	3313	22	AAU30134	Novel human secret
882	6	3.7	4630	18	AAW19629	Streptomyces venez
883	6	3.7	4630	21	AAY77177	S. venezuelae vep
884	6	3.7	4725	22	ABG23837	Novel human diagno
885	6	3.7	4977	22	ABG17057	Novel human diagno
886	6	3.7	6025	22	AAG84939	Shrimp white spot
887	6	3.7	6239	21	AAB23750	S. avermitilis ave
888	6	3.7	6239	22	AAG65265	Streptomyces averm
889	6	3.7	6685	22	ABG23030	Novel human diagno
890	6	3.7	7201	22	ABB71136	Drosophila melanog
891	6	3.7	10182	23	ABP38314	Staphylococcus epi
892	5	3.1	7	20	AAW92343	Latent fluorescent
893	5	3.1	8	19	AAW57875	C. felis esterase
894	5	3.1	8	22	AAE12921	Ctenocephalides fe
895	5	3.1	9	21	AAY97234	Complementary dete
896	5	3.1	9	22	AAE13142	Humanised antibody
897	5	3.1	9	22	AAG79232	Peptide of 45 kDa
898	5	3.1	9	22	AAB82708	VEGF antagonist an
899	5	3.1	9	22	AAB47113	Mutant scFv 508F l
900	5	3.1	9	22	AAG63993	Complementarity de
901	5	3.1	9	22	AAU15039	Antiangiogenic pep
902	5	3.1	9	22	AAU15040	Antiangiogenic pep
903	5	3.1	9	23	AAU95918	Immunogenic peptid
904	5	3.1	9	23	AAU74411	Light chain comple
905	5	3.1	9	23	AAE17068	Measles virus V pr
906	5	3.1	10	18	AAW31847	Release factor eRF
907	5	3.1	10	22	AAG93902	Human complementar
908	5	3.1	10	22	AAG94138	Human complementar
909	5	3.1	10	22	AAG97006	Human complementar
910	5	3.1	10	22	AAG97138	Human complementar
911	5	3.1	10	22	AAG84202	Arabidopsis thalia
912	5	3.1	10	22	AAG85675	Saccharomyces cere
913	5	3.1	10	22	AAG85677	Saccharomyces cere
914	5	3.1	10	22	AAB88256	Hsp-65 peptide epi
915	5	3.1	10	23	ABP61514	Human KRPI tryptic
916	5	3.1	11	21	AAY66797	T cell antigen rec
917	5	3.1	11	22	AAU25736	Breast cancer-asso
918	5	3.1	11	22	AAB98266	Anti-A33 antigen i
919	5	3.1	11	22	AAB75186	A33 antigen bindin
920	5	3.1	11	23	ABG67826	Human ADPI tryptic
921	5	3.1	12	19	AAW57871	C. felis esterase
922	5	3.1	12	19	AAW57873	C. felis esterase
923	5	3.1	12	22	ABB50920	Human secreted pro
924	5	3.1	12	22	AAE12917	Ctenocephalides fe
925	5	3.1	12	22	AAU68029	Human Breast cance
926	5	3.1	12	22	AAU68054	Human Breast cance
927	5	3.1	12	22	AAU68105	Human Breast cance
928	5	3.1	12	22	AAB72249	Colostrinin derive
929	5	3.1	12	22	AAB72503	Colostrinin peptid
930	5	3.1	12	22	AAB72535	Colostrinin peptid
931	5	3.1	12	22	AAB59323	Ewe colostrinin pe
932	5	3.1	12	23	AAE20231	Colostrinin consti
933	5	3.1	12	23	AAM51039	Colostrinin consti
934	5	3.1	12	23	AAO14580	Neural cell regula
935	5	3.1	12	23	AAE17060	Measles virus V pr
936	5	3.1	13	9	AAP81289	Atrial natriuretic
937	5	3.1	13	15	AAR46994	Apolipoprotein B-l
938	5	3.1	13	20	AAY34185	Human pre-prohepar
939	5	3.1	13	21	AAB39972	Anti-hIL12 antibod
940	5	3.1	13	21	AAB22033	Erythrocyte protei
941	5	3.1	13	21	AAB22078	Erythrocyte protei
942	5	3.1	13	21	AAY95875	O-glycosylation si
943	5	3.1	13	22	AAB99479	Erythrocyte band 4
944	5	3.1	13	22	AAB80876	Angiostatin-derive
945	5	3.1	13	22	AAB74270	Exemplary anti-ang
946	5	3.1	13	22	AAB85965	E. coli DNA polyme
947	5	3.1	13	22	AAB85968	T. maritima DNA po
948	5	3.1	13	22	AAB85971	T. neapolitana DNA
949	5	3.1	13	23	AAO18010	Human immunoglobul
950	5	3.1	13	23	AAU75514	Erythrocyte band 4
951	5	3.1	14	18	AAW31189	Human p62 internal
952	5	3.1	14	21	AAY99015	HLA class II bindi
953	5	3.1	14	22	AAM96786	Human peptide #61
954	5	3.1	14	22	AAM96895	Human peptide #170
955	5	3.1	14	22	AAM97438	Human peptide #713
956	5	3.1	14	22	AAM98587	Human peptide #186
957	5	3.1	14	22	AAM00679	Human protein frag
958	5	3.1	14	22	AAG98016	Human SNP associat
959	5	3.1	14	22	AAG98117	Human SNP associat

960	5	3.1	14	22	AAB59353	Ewe colostrinin pe
961	5	3.1	15	17	AAR94846	Peptide from libra
962	5	3.1	15	17	AAR94847	Peptide from libra
963	5	3.1	15	18	AAW43525	Mycobacteria sp. h
964	5	3.1	15	18	AAW43524	Mycobacteria sp. h
965	5	3.1	15	18	AAW23522	Purified cis-9,10-
966	5	3.1	15	19	AAY20515	Human neurofilamen
967	5	3.1	15	19	AAW57801	Fatty acid amide h
968	5	3.1	15	19	AAW62158	Agrobacterium faec
969	5	3.1	15	19	AAW62159	Agrobacterium faec
970	5	3.1	15	19	AAW62167	Agrobacterium faec
971	5	3.1	15	20	AAY43168	Integrin cell surf
972	5	3.1	15	21	AAB29733	Ganglioside GM1-bi
973	5	3.1	15	21	AAY92952	Transforming growt
974	5	3.1	15	21	AAY93104	Transforming growt
975	5	3.1	15	21	AAY66922	T cell antigen rec
976	5	3.1	15	21	AAY51055	EcoDNAPolI derived
977	5	3.1	15	21	AAY65488	Estradiol activate
978	5	3.1	15	22	AAB88294	Hsp-65 peptide epi
979	5	3.1	15	23	ABB09948	Integrin beta_6 su
980	5	3.1	15	23	AAO18022	Human immunoglobul
981	5	3.1	15	23	ABJ00523	Human IgE cyclic i
982	5	3.1	15	23	AAU86212	Oestradiol activat
983	5	3.1	15	23	ABB06199	Human DNA methylat
984	5	3.1	15	23	AAU73468	Vascular endotheli
985	5	3.1	15	23	ABB04314	Human Pax protein
986	5	3.1	16	21	AAB11423	Coxsackievirus B C
987	5	3.1	16	22	AAB88293	Hsp-65 peptide epi
988	5	3.1	16	23	AAE24517	Human RATL1d6 N-my
989	5	3.1	16	23	ABG67867	Human ADPI tryptic
990	5	3.1	16	23	AAE17067	Measles virus V pr
991	5	3.1	17	17	AAW01973	Rat pro-opiomelano
992	5	3.1	18	21	AAB22069	p67SRF protein fra
993	5	3.1	18	21	AAB22076	p67SRF protein fra
994	5	3.1	18	21	AAY95867	O-glycosylation si
995	5	3.1	18	22	AAB99477	p67-SRF peptide #3
996	5	3.1	18	22	AAB88292	Hsp-65 peptide epi
997	5	3.1	18	23	AAU75512	Transcription fact
998	5	3.1	19	17	AAR94848	Peptide from libra
999	5	3.1	19	18	AAW43526	Mycobacteria sp. h
1000	5	3.1	19	21	AAB51540	YadA homologous pe

ALIGNMENTS

```
RESULT 1
AAE18419
ID   AAE18419 standard; Protein; 279 AA.
XX
AC   AAE18419;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B240 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B240.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S, Revel A, Auguste C, Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29358.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B240 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             100.0%;  Score 161;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-156;
  Matches  161;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 2
ABG68154
ID   ABG68154 standard; Protein; 268 AA.
XX
AC   ABG68154;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Escherichia coli major fimbrial protein FimH.
XX
KW   Immunogen; FimH adhesin functional site-derived target peptide;
KW   FAFSD; helper T cell epitope; FimH; urinary tract infection;
KW   type 1 fimbriated uropathogenic enterobacteria; vaccine;
KW   FAFSD site-specific immunity.
XX
OS   Escherichia coli.
XX
PN   WO200251860-A2.
XX
PD   04-JUL-2002.
XX
PF   21-DEC-2001; 2001WO-US50816.
XX
PR   22-DEC-2000; 2000US-0747802.
XX
PA   (UNBI-) UNITED BIOMEDICAL INC.
XX
PI   Wang CY;
XX
DR   WPI; 2002-528681/56.
```

XX
PT Novel peptide immunogen, useful for evoking antibodies to prevent
PT adherence of Escherichia coli to bladder mucosa, comprises a FimH
PT adhesin functional site-derived target peptide covalently linked to
PT helper T cell epitope -
XX
PS Disclosure; Fig 1; 62pp; English.
XX
CC The invention describes a peptide immunogen (I), comprising a helper T
CC cell epitope sequence (Th) or a carrier protein covalently attached to a
CC FimH adhesin functional site-derived (FAFSD) target peptide comprising
CC not more than 30 amino acids of the carbohydrate binding pocket of FimH,
CC or its crossreactive and immunologically functional analogue or mimetope.
CC (I) and a composition containing (I) are useful for inducing anti-FAFSD
CC peptide antibody production in a mammal. The composition is also useful
CC for reducing adherence to the urinary tract mucosa of a mammal by type 1
CC fimbriated uropathogenic enterobacteria (Escherichia coli) to prevent
CC urinary tract infection. (I) has a focused FAFSD site-specific immunity
CC together with a broad protective immunity, and with less adverse side
CC reactions than the more complex polypeptide subunit vaccines and the
CC carrier conjugated vaccine. Since (I) is chemically well defined it is
CC easy and less costly to manufacture and to control or assure the quality
CC of the product. This is the amino acid sequence of FimH adhesin, peptides
CC derived from the functional site are used in the creation of a vaccine
CC against urinary tract infection described in the invention.
XX
SQ Sequence 268 AA;

Query Match 57.1%; Score 92; DB 23; Length 268;
Best Local Similarity 100.0%; Pred. No. 7.1e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy 61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
||||||||||||||||||||||||||||||||
Db 86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117

RESULT 3
AAB47074
ID AAB47074 standard; Protein; 279 AA.
XX
AC AAB47074;
XX
DT 08-MAY-2001 (first entry)
XX
DE Adhesin protein, FimH.
XX
KW Donor; pilus protein; pilin; adhesin; vaccine; urinary epithelia;
KW urinary tract infection; enterobacteriaceae.
XX
OS Escherichia coli.
XX
PN WO200104148-A2.
XX
PD 18-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-US19066.
XX
PR 13-JUL-1999; 99US-0143582.
PR 16-JUL-1999; 99US-0144359.
PR 23-FEB-2000; 2000US-0184442.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Hultgren SJ, Pinkner JS, Sauer F, Barnhart M, Waksman G, Knight S;
XX
DR WPI; 2001-138315/14.
XX
PT Immunogenic complexes and polypeptides for vaccinating against urinary
PT tract disease, comprises a pilus protein component and a bacterial
PT chaperone -
XX
PS Disclosure; Fig 1; 92pp; English.
XX
CC This sequence may be used as the pilus protein in the immunogenic
CC complex of the invention. The complex comprises a pilus protein
CC component and a donor complement portion as part of the same amino
CC acid sequence or as non-covalently linked fragments of a complex
CC such that the correct conformation of the pilin is maintained. The
CC pilus protein component may be an adhesin or a pilin. Pilus associated
CC adhesins, such as FimH are relatively conserved proteins among
CC different species and strains of bacteria, therefore vaccines
CC incorporating the FimH antigen exhibit a broad spectrum of
CC protection compared with current pilus-fiber based vaccines. The
CC immunogenic complexes act by disrupting pilus-mediated attachment
CC of E. coli to urinary epithelia and may prevent or retard the
CC development of urinary tract infections. Vaccines containing the
CC complexes are useful for preventing urinary tract disease in a human
CC caused by the bacterium family enterobacteriaceae specifically
CC Escherichia coli and may also be used in treating the disease.
XX
SQ Sequence 279 AA;

Query Match 57.1%; Score 92; DB 22; Length 279;
Best Local Similarity 100.0%; Pred. No. 7.4e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy 61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
||||||||||||||||||||||||||||||||
Db 86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117

RESULT 4
AAB72839
ID AAB72839 standard; Protein; 279 AA.
XX
AC AAB72839;
XX
DT 09-MAY-2001 (first entry)
XX
DE Bacterial type 1 pilus subunit FimA.
XX
KW Antibacterial compound; Gram-negative bacterium; pilus; chaperone;
KW biofilm; disease treatment; bacterial infection.
XX
OS Bacteria.
XX
PN WO200110386-A2.
XX
PD 15-FEB-2001.
XX
PF 11-AUG-2000; 2000WO-US22087.
XX
PR 11-AUG-1999; 99US-0148280.
XX
PA (UNIW) UNIV WASHINGTON.
XX
PI Hultgren SJ, Sauer FG, Waksman G, Fuetterer K;
XX
DR WPI; 2001-226496/23.
XX
PT An isolated compound for inhibiting pilus assembly -
XX
PS Example 5; Fig 8A; 144pp; English.
XX
CC The present invention provides antibacterial compounds which are able to
CC interfere with Gram-negative bacteria pilus formation and assembly, and
CC pilus interaction with chaperone proteins. These are useful in the

```
CC   treatment of bacterial infection, and in the prevention of biofilm
CC   formation. They are particularly useful against Escherichia coli,
CC   Haemophilus influenzae, Salmonella enteritidis, S. typhimurium,
CC   Bordetella pertussis, Yersinia enterocolitica, Helicobacter pylori and
CC   Klebsiella pneumoniae.
XX
SQ   Sequence   279 AA;

  Query Match          57.1%;  Score 92;  DB 22;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-86;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 5
AAY72515
ID   AAY72515 standard; Protein; 279 AA.
XX
AC   AAY72515;
XX
DT   02-MAY-2001  (first entry)
XX
DE   Escherichia coli FimH protein.
XX
KW   FimH; adhesin protein; type 1 pilus; mannose binding domain; MBD; COL;
KW   collagen binding domain; prophylaxis; therapy; urinary tract infection;
KW   UTI; immunogen; passive immunotherapy; vaccine; antibacterial.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..156
FT                   /label= Lectin_binding_domain
FT   Domain          1..20
FT                   /label= MBD-1
FT                   /note= "Mannose binding domain-1"
FT   Region          2..5
FT                   /note= "Beta strand 1"
FT   Region          9..11
FT                   /note= "Beta strand 2"
FT   Region          17..25
FT                   /note= "Beta strand 3"
FT   Binding-site    27
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    31
FT                   /label= FimC_chaperone_binding_site
FT   Region          32..37
FT                   /note= "Beta strand 4a"
FT   Region          38..41
FT                   /note= "3-10 helix"
FT   Region          42..46
FT                   /note= "Beta strand 4b"
FT   Domain          46..54
FT                   /label= MBD-2
FT                   /note= "Mannose binding domain 2"
FT   Domain          50..80
FT                   /label= COL
FT                   /note= "Collagen binding domain;  Forms a beta-sheet-
FT                   alpha-1-beta-sheet structure from strands 5, alpha-1
FT                   and 6"
FT   Region          54..63
FT                   /note= "Beta strand 5"
FT   Region          67..69
FT                   /note= "Alpha-1 helix"
FT   Region          71..77
FT                   /note= "Beta strand 6"
FT   Region          80..83
FT                   /note= "Beta strand 7"
FT   Region          89..95
FT                   /note= "Beta strand 8"
FT   Region          104..111
FT                   /note= "Beta strand 9"
FT   Binding-site    117
FT                   /label= FimC_chaperone_binding_site
FT   Region          125..135
FT                   /note= "Beta strand 10"
FT   Domain          127..148
FT                   /label= MBD-3
FT                   /note= "Mannose binding domain 3"
FT   Region          139..150
FT                   /note= "Beta strand 11"
FT   Binding-site    155
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    157..158
FT                   /label= FimC_chaperone_binding_site
FT   Domain          159..279
FT                   /label= Pilin_domain
FT   Binding-site    160
FT                   /label= FimC_chaperone_binding_site
FT   Region          161..165
FT                   /note= "Beta strand A'"
FT   Binding-site    162..163
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    168..171
FT                   /note= "FimC_chaperone binding site; Serves as beta
FT                   strand A''"
FT   Binding-site    173
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    181
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    183
FT                   /label= FimC_chaperone_binding_site
FT   Region          183..187
FT                   /note= "Beta strand B"
FT   Binding-site    191..192
FT                   /label= FimC_chaperone_binding_site
FT   Region          192..197
FT                   /note= "Beta strand C"
FT   Binding-site    196
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    198
FT                   /label= FimC_chaperone_binding_site
FT   Binding-site    215
FT                   /label= FimC_chaperone_binding_site
FT   Region          221..226
FT                   /note= "Beta strand D'"
FT   Region          235..239
FT                   /note= "Beta strand D''"
FT   Region          251..258
FT                   /note= "Beta strand E"
FT   Binding-site    266..279
FT                   /label= FimC_chaperone_binding_site
FT   Region          268..279
FT                   /note= "Beta strand F; contains conserved beta-
FT                   zipper motif"
XX
PN   WO200105978-A1.
XX
PD   25-JAN-2001.
XX
PF   14-JUL-2000; 2000WO-US19402.
XX
PR   15-JUL-1999;   99US-0144016.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Hultgren SJ,  Langermann S;
XX
DR   WPI; 2001-159539/16.
```

```
XX
PT   Polypeptides useful as vaccines for prevention and/or treatment of
PT   diseases such as urinary tract infections, caused by
PT   Enterobacteriaceae, comprises mannose-binding domains derived from
PT   adhesin molecules -
XX
PS   Claim 1; Fig 3; 53pp; English.
XX
CC   The present sequence is FimH protein from Escherichia coli. FimH
CC   is an adhesin protein found in type 1 pili of bacteria of the
CC   family enterobacteriaceae, especially E. coli. The FimH
CC   protein comprises mannose-binding domains (MBDs) and collagen-
CC   binding domains (COL). The present invention relates to
CC   engineered polypeptides comprising one or more domains derived
CC   from FimH protein. These polypeptides are used to produce
CC   prophylactic vaccines which are useful for the prevention
CC   and/or treatment of diseases, such as urinary tract infection (UTI)
CC   caused by a bacterium of the family Enterobacteriaceae, especially
CC   E. coli in animals, in particular humans. They are useful as immunogens
CC   to stimulate the production of antibodies for use in passive immuno-
CC   therapy, as a diagnostic reagent and as a reagent in other processes
CC   such as affinity chromatography. The antibodies of the novel poly-
CC   peptides are also useful for research purposes for studying protein-
CC   lectin or collagen binding interactions.
XX
SQ   Sequence   279 AA;

  Query Match             57.1%;  Score 92;  DB 22;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-86;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 6
AAE18424
ID   AAE18424 standard; Protein; 279 AA.
XX
AC   AAE18424;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC56 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC56.
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29363.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC56 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             57.1%;  Score 92;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-86;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 7
AAE18426
ID   AAE18426 standard; Protein; 279 AA.
XX
AC   AAE18426;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC60 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC60.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 27
FT                   /note= "Encoded by GCC"
FT   Misc-difference 66
FT                   /note= "Encoded by AGC"
FT   Misc-difference 70
FT                   /note= "Encoded by AGT"
FT   Misc-difference 78
FT                   /note= "Encoded by AAT"
FT   Misc-difference 119
FT                   /note= "Encoded by GCG"
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 269
FT                   /note= "Encoded by CAA"
FT   Misc-difference 273
FT                   /note= "Encoded by GGC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
```

XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29365.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC60 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 57.1%; Score 92; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 7.4e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117
```

RESULT 8
AAE18433
ID AAE18433 standard; Protein; 279 AA.
XX
AC AAE18433;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain J96 FimH protein.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder.
XX
OS Escherichia coli J96.
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR N-PSDB; AAD29372.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Claim 3; Page 90-91; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain J96 FimH protein.
XX
SQ Sequence 279 AA;

Query Match 57.1%; Score 92; DB 23; Length 279;
Best Local Similarity 100.0%; Pred. No. 7.4e-86;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117
```

RESULT 9
AAB47073
ID AAB47073 standard; Protein; 296 AA.
XX
AC AAB47073;
XX
DT 08-MAY-2001 (first entry)
XX
DE Immunogenic complex: FimH-linker-FimG N-terminal extension.
XX
KW Donor; pilus protein; pilin; adhesin; vaccine; urinary epithelia;
KW urinary tract infection; enterobacteriaceae.
XX
OS Chimeric - Escherichia coli.
OS Chimeric - Synthetic.
XX
FH Key Location/Qualifiers
FT Protein 1..279
FT /label= FimH
FT /note= "Pilus protein component"
FT Peptide 280..283
FT /note= "Linker"
FT Peptide 284..296
FT /label= N-terminal extension of FimG
FT /note= "Donor strand component"
XX
PN WO200104148-A2.
XX
PD 18-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-US19066.
XX
PR 13-JUL-1999; 99US-0143582.
PR 16-JUL-1999; 99US-0144359.
PR 23-FEB-2000; 2000US-0184442.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Hultgren SJ, Pinkner JS, Sauer F, Barnhart M, Waksman G, Knight S;
XX
DR WPI; 2001-138315/14.
XX
PT Immunogenic complexes and polypeptides for vaccinating against urinary
PT tract disease, comprises a pilus protein component and a bacterial
PT chaperone -
XX

```
PS   Claim 19; Page 81-82; 92pp; English.
XX
CC   This sequence represents the immunogenic complex of the invention.
CC   The complex comprises a pilus protein component and a donor complement
CC   portion as part of the same amino acid sequence attached through an
CC   amino acid linker. The linker is composed of a sequence which
CC   readily forms a loop such that the donor strand can loop back towards
CC   the pilus protein and form an anti-parallel structure.  Pilus associated
CC   adhesins, such as FimH are relatively conserved proteins among
CC   different species and strains of bacteria, therefore vaccines
CC   incorporating the FimH antigen exhibit a broad spectrum of
CC   protection compared with current pilus-fiber based vaccines.  The
CC   immunogenic complexes act by disrupting pilus-mediated attachment
CC   of E. coli to urinary epithelia and may prevent or retard the
CC   development of urinary tract infections.  Vaccines containing the
CC   complexes are useful for preventing urinary tract disease in a human
CC   caused by the bacterium family enterobacteriaceae specifically
CC   Escherichia coli and may also be used in treating the disease.
XX
SQ   Sequence   296 AA;

  Query Match             57.1%;  Score 92;  DB 22;  Length 296;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-86;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 10
AAR76745
ID   AAR76745 standard; protein; 300 AA.
XX
AC   AAR76745;
XX
DT   13-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli K12 strain PC31.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli K12 strain PC31.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 88-89; 152pp; English.
XX
CC   This sequence represents the FimH protein from E. coli K12 strain
CC   PC31.  FimH is located at the tip of the  type 1 fimbriae and also
CC   intercalated at intervals in the fimbrial organelle.  Most forms of
CC   the FimH adhesin target, and bind to, oligosaccharide structures
CC   containing terminally located alpha-D-mannoside residues.  FimH
CC   contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  This
CC   sequence and those given in AAR76763-76 may be used in the production of
CC   a variant FimH adhesin which may be useful for targetting active
CC   compounds and microbial cells to locations comprising selected receptors
CC   to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             57.1%;  Score 92;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-86;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 11
AAB16009
ID   AAB16009 standard; Protein; 300 AA.
XX
AC   AAB16009;
XX
DT   05-OCT-2000  (first entry)
XX
DE   E. coli proliferation associated protein sequence SEQ ID NO:367.
XX
KW   Escherichia coli; E. coli; proliferation; inhibition; screening;
KW   antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
OS   Escherichia coli.
XX
PN   WO200044906-A2.
XX
PD   03-AUG-2000.
XX
PF   27-JAN-2000; 2000WO-US02200.
XX
PR   27-JAN-1999;   99US-0117405.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zyskind J,  Ohlsen KL,  Trawick J,  Forsyth RA,  Froelich JM,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2000-514822/46.
DR   N-PSDB; AAA66015.
XX
PT   Novel polynucleotides and polypeptides associated with microorganism
PT   proliferation, used to identify inhibitors of bacterial growth and
PT   proliferation, for use in antisense therapy -
XX
PS   Claim 11; Page 274-275; 316pp; English.
```

```
XX
CC   AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC   sequences derived from Escherichia coli which inhibit E. coli
CC   proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC   nucleotide and protein sequences associated with E. coli proliferation.
CC   AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC   proliferation inhibiting nucleotide inserts in an example from the
CC   present invention. Methods from the present invention can be used to
CC   identify a proliferation- required gene in a microorganism, by contacting
CC   a microorganism with a proliferation-required gene activity inhibitory
CC   nucleic acid identified in another organism, and determining if
CC   inhibition occurs in the second microorganism. The nucleic acid sequences
CC   identified as being required for bacterial growth and proliferation, can
CC   be used for antisense therapy for killing bacteria.
XX
SQ   Sequence   300 AA;

  Query Match             57.1%;  Score 92;  DB 21;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-86;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 12
AAY59456
ID   AAY59456 standard; peptide; 300 AA.
XX
AC   AAY59456;
XX
DT   29-MAR-2000  (first entry)
XX
DE   E. coli PC31 FimH protein.
XX
KW   Multifunctional adhesin protein; organic receptor; bioremediation;
KW   biosorption; organic pollutant; herbicide; pesticide; toxic compound;
KW   recycling; metal isolation; metal binding domain.
XX
OS   Escherichia coli.
XX
PN   WO9957276-A1.
XX
PD   11-NOV-1999.
XX
PF   21-APR-1999;   99WO-DK00223.
XX
PR   30-APR-1998;   98DK-0000598.
PR   01-MAY-1998;   98US-0083794.
XX
PA   (GYRE-) GYRE LTD.
XX
PI   Schembri MA,  Klemm P;
XX
DR   WPI; 2000-072233/06.
XX
PT   Novel recombinant cells useful for bioremediation and recycling
PT   processes -
XX
PS   Claim 5; Page 5; 60pp; English.
XX
CC   This sequence is the E. coli FimH protein, which is an adhesin protein.
CC   The invention relates to a recombinant cell expressing a multifunctional
CC   adhesin (MA) protein on its surface. The MA protein has at least one
CC   binding domain (BD1) capable of binding to an organic receptor, and at
CC   least one binding domain (BD2) not naturally present in the adhesin, and
CC   can bind to a compound to which the naturally occurring adhesin protein
CC   does not substantially bind. Cells of the invention are used as
CC   bioremediation or biosorption means to separate undesired compounds such
CC   as organic pollutants including herbicides and pesticides, or toxic
CC   compounds such as heavy metals from the environment, or for isolating
CC   precious compounds such as metals for recycling purposes.
XX
SQ   Sequence   300 AA;

  Query Match             57.1%;  Score 92;  DB 21;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-86;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 13
AAU77488
ID   AAU77488 standard; Protein; 300 AA.
XX
AC   AAU77488;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Escherichia coli type I pilin protein, FimH.
XX
KW   Immune response; primate; immunoglobulin; urogenital tract infection;
KW   FimC-FimH complex; FimCH; IgG; human; urinary tract infection; UTI;
KW   bladder infection; kidney infection; Enterobacteriaceae; bactourea;
KW   pregnant woman; diabetic; immunocompromised; HIV; cancer;
KW   human immunodeficiency virus infection; end stage renal disease;
KW   type I pilin protein; adhesin; FimH.
XX
OS   Escherichia coli isolate J96.
XX
PN   WO200215928-A1.
XX
PD   28-FEB-2002.
XX
PF   28-NOV-2000; 2000WO-US32398.
XX
PR   18-AUG-2000; 2000US-226146P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Ballou WR;
XX
DR   WPI; 2002-280859/32.
DR   N-PSDB; ABK11187.
XX
PT   Stimulating immune response in a primate for preventing, treating
PT   bacterial induced diseases such as diseases of urinary tract, by
PT   administering bacterial adhesive proteins, preferably FimC-FimH
PT   polypeptide complex -
XX
PS   Claim 8; Page 89-90; 92pp; English.
XX
CC   The present invention relates to a method of inducing an immune response
CC   in a primate. The response involves immunoglobulin (Ig) molecules that
CC   bind a bacterial adhesin protein, preferably an attachment domain of
CC   a type 1 pilin polypeptide (e.g. FimH) associated with a bacterium
CC   causing urogenital tract infections (e.g. Escherichia coli). The method
CC   comprises administering a purified FimH polypeptide, a FimC-FimH (FimCH)
CC   complex, or immunogenic fragments of these. The method is useful for
CC   inducing IgG molecules in a primate, especially human, to reduce or
CC   prevent the incidence of urogenital tract infections, particularly
CC   urinary tract infection (UTI), bladder infection, or kidney infection,
CC   caused by a bacterium of the family Enterobacteriaceae, preferably
CC   E. coli. The method can be used in a human subject that has suffered
```

```
CC   more than two urogenital infections within one year, has asymptomatic
CC   bactourea, is a pregnant woman or a diabetic, is immunocompromised, has
CC   a human immunodeficiency virus (HIV) infection, has cancer, is in
CC   remission from cancer, or is at risk for end stage renal disease. The
CC   method is useful for vaccinating a primate against urogenital tract
CC   infections, for treating or ameliorating the symptoms of urogenital
CC   tract infections, and also for slowing or preventing progression of
CC   a urinary tract infection into end stage renal disease. The present
CC   sequence represents E. coli FimH protein.
XX
SQ   Sequence   300 AA;

  Query Match           57.1%;  Score 92;  DB 23;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-86;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 14
AAB47072
ID   AAB47072 standard; Protein; 304 AA.
XX
AC   AAB47072;
XX
DT   08-MAY-2001  (first entry)
XX
DE   Immunogenic complex: FimH-linker-G1 beta-strand of FimC.
XX
KW   Donor; pilus protein; pilin; adhesin; vaccine; urinary epithelia;
KW   urinary tract infection; enterobacteriaceae.
XX
OS   Chimeric - Escherichia coli.
OS   Chimeric - Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..279
FT                   /label= FimH
FT                   /note= "Pilus protein component"
FT   Peptide         280..289
FT                   /note= "Linker"
FT   Peptide         290..304
FT                   /label= G1 beta-strand of FimC
FT                   /note= "Donor strand component"
XX
PN   WO200104148-A2.
XX
PD   18-JAN-2001.
XX
PF   13-JUL-2000; 2000WO-US19066.
XX
PR   13-JUL-1999;   99US-0143582.
PR   16-JUL-1999;   99US-0144359.
PR   23-FEB-2000; 2000US-0184442.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Hultgren SJ,  Pinkner JS,  Sauer F,  Barnhart M,  Waksman G,  Knight S;
XX
DR   WPI; 2001-138315/14.
XX
PT   Immunogenic complexes and polypeptides for vaccinating against urinary
PT   tract disease, comprises a pilus protein component and a bacterial
PT   chaperone -
XX
PS   Claim 19; Page 80-81; 92pp; English.
XX
CC   This sequence represents the immunogenic complex of the invention.
CC   The complex comprises a pilus protein component and a donor complement
CC   portion as part of the same amino acid sequence attached through an
CC   amino acid linker. The linker is composed of a sequence which
CC   readily forms a loop such that the donor strand can loop back towards
CC   the pilus protein and form an anti-parallel structure.  Pilus associated
CC   adhesins, such as FimH are relatively conserved proteins among
CC   different species and strains of bacteria, therefore vaccines
CC   incorporating the FimH antigen exhibit a broad spectrum of
CC   protection compared with current pilus-fiber based vaccines.  The
CC   immunogenic complexes act by disrupting pilus-mediated attachment
CC   of E. coli to urinary epithelia and may prevent or retard the
CC   development of urinary tract infections.  Vaccines containing the
CC   complexes are useful for preventing urinary tract disease in a human
CC   caused by the bacterium family enterobacteriaceae specifically
CC   Escherichia coli and may also be used in treating the disease.
XX
SQ   Sequence   304 AA;

  Query Match           57.1%;  Score 92;  DB 22;  Length 304;
  Best Local Similarity   100.0%;  Pred. No. 8e-86;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 15
AAE18417
ID   AAE18417 standard; Protein; 279 AA.
XX
AC   AAE18417;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B228 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B228.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 203
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29356.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
```

```
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B228 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             55.9%;  Score 90;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           |||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 16
AAE18418
ID   AAE18418 standard; Protein; 279 AA.
XX
AC   AAE18418;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B238 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B238.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 24
FT                   /note= "Encoded by ATT"
FT   Misc-difference 26
FT                   /note= "Encoded by CCC"
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 273
FT                   /note= "Encoded by GCC"
FT   Misc-difference 274
FT                   /note= "Encoded by GTG"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29357.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B238 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             55.9%;  Score 90;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           |||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 17
AAE18422
ID   AAE18422 standard; Protein; 279 AA.
XX
AC   AAE18422;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC42 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC42.
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29361.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
```

```
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC42 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             55.9%;  Score 90;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy     63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||
Db     88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 18
AAE18429
ID   AAE18429 standard; Protein; 279 AA.
XX
AC   AAE18429;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC80 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC80.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 27
FT                   /note= "Encoded by GTC"
FT   Misc-difference 165
FT                   /note= "Encoded by GCT"
FT   Misc-difference 171
FT                   /note= "Encoded by ACT"
FT   Misc-difference 269
FT                   /note= "Encoded by CAA"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29368.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC80 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             55.9%;  Score 90;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy     63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||
Db     88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 19
AAE18432
ID   AAE18432 standard; Protein; 279 AA.
XX
AC   AAE18432;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain G189 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli G189.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29371.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain G189 FimH protein.
XX
SQ   Sequence   279 AA;
```

```
  Query Match             55.9%;  Score 90;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 20
AAE18436
ID   AAE18436 standard; Protein; 279 AA.
XX
AC   AAE18436;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli FimH consensus protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli.
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli FimH consensus protein.
XX
SQ   Sequence   279 AA;

  Query Match             55.9%;  Score 90;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 21
AAR76769
ID   AAR76769 standard; protein; 300 AA.
XX
AC   AAR76769;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CI#3.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CI#3.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93071.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             55.9%;  Score 90;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
```

```
                ||||||||||||||||||||||||||||||
Db    109 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 22
AAR76771
ID   AAR76771 standard; protein; 300 AA.
XX
AC   AAR76771;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate KS-54.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate KS-54.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93073.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             55.9%;  Score 90;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy     63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||
Db    109 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 23
AAR76773
ID   AAR76773 standard; protein; 300 AA.
XX
AC   AAR76773;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate MJ#9-3.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate MJ#9-3.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93067.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             55.9%;  Score 90;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db     109 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 24
AAR76774
ID   AAR76774 standard; protein; 300 AA.
XX
AC   AAR76774;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate MJ#31-3.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate MJ#31-3.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93068.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;
```

```
  Query Match             55.9%;  Score 90;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db     109 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 25
AAR76763
ID   AAR76763 standard; protein; 300 AA.
XX
AC   AAR76763;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate KB21.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate KB21.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
```

```
SQ   Sequence   300 AA;

  Query Match             55.9%;  Score 90;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-84;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db     109 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 26
AAR76767
ID   AAR76767 standard; protein; 296 AA.
XX
AC   AAR76767;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CI#10.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CI#10.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..296
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93069.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   296 AA;

  Query Match             55.3%;  Score 89;  DB 16;  Length 296;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-83;
  Matches   89;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAG 91
           |||||||||||||||||||||||||||||
Db     109 SETPRVVYNSRTDKPWPVALYLTPVSSAG 137


RESULT 27
AAR76772
ID   AAR76772 standard; protein; 300 AA.
XX
AC   AAR76772;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate U221-3.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate U221-3.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93074.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
```

```
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             49.1%;  Score 79;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-72;
  Matches   79;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy     63 SETPRVVYNSRTDKPWPVA 81
          ||||||||||||||||||||
Db    109 SETPRVVYNSRTDKPWPVA 127


RESULT 28
AAE18420
ID   AAE18420 standard; Protein; 280 AA.
XX
AC   AAE18420;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B242 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B242.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 279..280
FT                   /note= "Encoded by CAA"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29359.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B242 FimH protein.
XX
SQ   Sequence   280 AA;

  Query Match             46.6%;  Score 75;  DB 23;  Length 280;
  Best Local Similarity   100.0%;  Pred. No. 2e-68;
  Matches   75;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     18 FCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKP 77
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     43 FCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKP 102

Qy     78 WPVALYLTPVSSAGG 92
          |||||||||||||||
Db    103 WPVALYLTPVSSAGG 117


RESULT 29
AAE18415
ID   AAE18415 standard; Protein; 279 AA.
XX
AC   AAE18415;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B217 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B217.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 62
FT                   /note= "Encoded by TCG"
FT   Misc-difference 70
FT                   /note= "Encoded by AAT"
FT   Misc-difference 78
FT                   /note= "Encoded by AGT"
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 234
FT                   /note= "Encoded by ACG"
FT   Misc-difference 244
FT                   /note= "Encoded by CGG"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29354.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
```

```
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B217 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             41.6%;  Score 67;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-60;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 30
AAE18416
ID   AAE18416 standard; Protein; 279 AA.
XX
AC   AAE18416;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B223 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B223.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29355.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B223 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             41.6%;  Score 67;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-60;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 31
AAE18425
ID   AAE18425 standard; Protein; 279 AA.
XX
AC   AAE18425;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC58 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC58.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29364.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC58 FimH protein.
XX
SQ   Sequence   279 AA;
```

```
  Query Match             41.6%;  Score 67;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-60;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 32
AAE18427
ID   AAE18427 standard; Protein; 279 AA.
XX
AC   AAE18427;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC61 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC61.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 3
FT                   /note= "Encoded by TAT"
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29366.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC61 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             41.6%;  Score 67;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-60;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 33
AAE18430
ID   AAE18430 standard; Protein; 279 AA.
XX
AC   AAE18430;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC89 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC89.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 226
FT                   /note= "Encoded by GCG"
FT   Misc-difference 227
FT                   /note= "Encoded by CGC"
FT   Misc-difference 231
FT                   /note= "Encoded by GTT"
FT   Misc-difference 232
FT                   /note= "Encoded by ATT"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29369.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC89 FimH protein.
XX
SQ   Sequence   279 AA;
```

```
  Query Match            41.6%;  Score 67;  DB 23;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-60;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 34
AAE18431
ID   AAE18431 standard; Protein; 279 AA.
XX
AC   AAE18431;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC95 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC95.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29370.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Page 87-88; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC95 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match            41.6%;  Score 67;  DB 23;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-60;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 35
AAE18434
ID   AAE18434 standard; Protein; 279 AA.
XX
AC   AAE18434;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain NU14 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli NU14.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29373.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain NU14 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match            41.6%;  Score 67;  DB 23;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-60;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186
```

```
Db    180 PIPLTVY 186


RESULT 36
AAE18435
ID   AAE18435 standard; Protein; 279 AA.
XX
AC   AAE18435;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain G162 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli G162.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29382.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Page 88-89; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain G162 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             41.6%;  Score 67;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-60;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 37
AAE18428
ID   AAE18428 standard; Protein; 279 AA.
XX
AC   AAE18428;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC62 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC62.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29367.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC62 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             37.3%;  Score 60;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-53;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     102 AVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127 AVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 38
AAR76768
ID   AAR76768 standard; protein; 300 AA.
XX
AC   AAR76768;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CSH-50.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
```

```
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CSH-50.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             36.6%;  Score 59;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-52;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      34 QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 39
AAR76766
ID   AAR76766 standard; protein; 300 AA.
XX
AC   AAR76766;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CI#7.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CI#7.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93072.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             35.4%;  Score 57;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 6e-50;
  Matches   57;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 LVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSET 65
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      55 LVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSET 111


RESULT 40
AAE18423
ID   AAE18423 standard; Protein; 279 AA.
XX
AC   AAE18423;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC45 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli EC45.
```

```
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29362.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC45 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             34.8%;  Score 56;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-49;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 175


RESULT 41
AAR76765
ID   AAR76765 standard; protein; 300 AA.
XX
AC   AAR76765;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CI#4.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CI#4.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93065.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             34.8%;  Score 56;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-49;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 196


RESULT 42
AAR76770
ID   AAR76770 standard; protein; 300 AA.
XX
AC   AAR76770;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate F-18.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate F-18.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
```

```
XX
PI   Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93063.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates. FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains. For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes. Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein. The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein. The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle. These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             34.8%;  Score 56;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-49;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 196


RESULT 43
ABB09458
ID   ABB09458 standard; Protein; 408 AA.
XX
AC   ABB09458;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Fusion protein prtY-FimH-prtY.
XX
KW   Exoproteinase; immunostimulant; vaccine; anchor peptide;
KW   FimH; ciliated adhesive factor.
XX
OS   Lactobacillus helveticus.
OS   Escherichia coli.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..70
FT                   /label= signal_peptide
FT                   /note= "derived from L. helveticus"
FT   Protein         71..295
FT                   /label= FimH_fragment
FT                   /note= "derived from E. coli ciliated adhesive factor"
FT   Peptide         296..408
FT                   /label= anchor_peptide
FT                   /note= "derived from L. helveticus"
XX
PN   JP2002017357-A.
XX
PD   22-JAN-2002.
XX
PF   04-JUL-2000; 2000JP-0202442.
XX
PR   04-JUL-2000; 2000JP-0202442.
XX
PA   (CALV ) CALPIS SHOKUHIN KOGYO KK.
XX
DR   WPI; 2002-221706/28.
DR   N-PSDB; ABL52756.
XX
PT   New anchor peptide, useful for anchoring protein onto microbe -
XX
PS   Example 1; Page 10-11; 12pp; Japanese.
XX
CC   The invention relates to an anchor peptide derived from the
CC   exoproteinase of Lactobacillus helveticus FERM BP-6060.
CC   The activity of compositions of the invention may be described as
CC   immunostimulatory. The invention also includes a fusion protein,
CC   containing the anchor peptide and a signal peptide fused to a useful
CC   protein. The invention also includes a method for anchoring a useful
CC   protein onto the surface of a microbe. The microbe can be used as a
CC   vaccine. The current sequence represents a fusion protein referred to as
CC   prtY-FimH-prtY. This fusion protein contains a fragment of the FimH
CC   protein, which is a ciliated adhesive factor from E. coli, arranged
CC   between the the signal sequence and anchor sequences derived from L.
CC   helveticus exoproteinase (see ABB09456 and ABB09457).
XX
SQ   Sequence   408 AA;

  Query Match             34.8%;  Score 56;  DB 23;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 8.5e-49;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     187 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 242


RESULT 44
AAE18413
ID   AAE18413 standard; Protein; 279 AA.
XX
AC   AAE18413;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B210 FimH protein.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder.
XX
OS   Escherichia coli B210.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 141
FT                   /note= "Encoded by GGT"
FT   Misc-difference 176
FT                   /note= "Encoded by CCT"
FT   Misc-difference 201
FT                   /note= "Encoded by ACC"
FT   Misc-difference 203
FT                   /note= "Encoded by GGT"
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
```

```
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   N-PSDB; AAD29352.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Claim 3; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B210 FimH protein.
XX
SQ   Sequence   279 AA;

  Query Match             29.8%;  Score 48;  DB 23;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-41;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG 48
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG 73


RESULT 45
AAR76764
ID   AAR76764 standard; protein; 300 AA.
XX
AC   AAR76764;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein derived from E. coli clinical isolate CI#12.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site.
XX
OS   Escherichia coli clinical isolate CI#12.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "Signal peptide"
FT   Protein         22..300
FT                   /note= "Mature FimH"
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   N-PSDB; AAQ93075.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 1; Page 44-45; 152pp; English.
XX
CC   The sequences given in AAR76763-76 are FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of a variant FimH adhesin which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence   300 AA;

  Query Match             28.6%;  Score 46;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-38;
  Matches   46;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSA 140
           ||||||||||||||||||||||||||||||||||||||||||||||
Db     141 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSA 186


Search completed: November 28, 2002, 19:03:02
Job time : 81 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 18:54:51 ; Search time 25 Seconds
(without alignments)
267.108 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 161
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	92	57.1	300	1	FIMH_ECOLI	P08191	escherichia
2	9	5.6	304	1	YDEQ_ECOLI	P77588	escherichia
3	7	4.3	120	1	RL34_ARATH	Q42351	arabidopsis
4	7	4.3	120	1	RL34_PEA	P40590	pisum sativ
5	7	4.3	120	1	RL34_TOBAC	P41098	nicotiana t
6	7	4.3	131	1	MYP2_HUMAN	P02689	homo sapien
7	7	4.3	291	1	YQHO_BACSU	P54513	bacillus su
8	7	4.3	338	1	YAS4_SCHPO	Q10140	schizosacch
9	7	4.3	629	1	CTR1_HUMAN	P30825	homo sapien
10	7	4.3	753	1	YBHJ_ECOLI	P75764	escherichia
11	7	4.3	1040	1	YEGN_ECOLI	P76398	escherichia
12	6	3.7	68	1	VPO_BPPRD	P27390	bacteriopha
13	6	3.7	85	1	IVB2_BUNMU	P00989	bungarus mu
14	6	3.7	111	1	YDGC_ECOLI	P52110	escherichia
15	6	3.7	122	1	RFA3_YEAST	P26755	saccharomyc
16	6	3.7	133	1	NB5M_CHICK	P48306	gallus gall
17	6	3.7	137	1	YEDX_ECO57	Q8xb75	escherichia
18	6	3.7	137	1	YEDX_ECOLI	P76341	escherichia
19	6	3.7	142	1	RS9_PYRAE	Q8zyq0	pyrobaculum
20	6	3.7	149	1	H2A2_PEA	P40281	pisum sativ
21	6	3.7	150	1	H2A1_PEA	P25470	pisum sativ
22	6	3.7	154	1	NAPB_RHOSH	Q53177	rhodobacter
23	6	3.7	159	1	LSPA_STACA	Q59835	staphylococ
24	6	3.7	178	1	RL6_BACSU	P46898	bacillus su
25	6	3.7	179	1	PTH_CHLMU	P49607	chlamydia m
26	6	3.7	194	1	RUVA_FUSNN	Q8rej7	fusobacteri
27	6	3.7	209	1	RL3_BUCAI	P57591	buchnera ap
28	6	3.7	213	1	MUC3_HUMAN	Q02505	homo sapien
29	6	3.7	215	1	RPB5_KLUMA	Q9p4b9	kluyveromyc
30	6	3.7	215	1	RPB5_YEAST	P20434	saccharomyc
31	6	3.7	216	1	ALKB_SALTY	P37462	salmonella
32	6	3.7	217	1	CAT_PROMI	P07641	proteus mir
33	6	3.7	217	1	R10A_SPOFR	Q963b6	spodoptera
34	6	3.7	217	1	VTI1_YEAST	Q04338	saccharomyc
35	6	3.7	218	1	GCH1_HAEIN	P43866	haemophilus
36	6	3.7	219	1	GLNP_ECOLI	P10345	escherichia
37	6	3.7	225	1	VV_PI2H	P19847	human parai
38	6	3.7	225	1	VV_PI2HT	P23057	human parai
39	6	3.7	231	1	TRYP_PIG	P00761	sus scrofa
40	6	3.7	235	1	SODM_MAIZE	P09233	zea mays (m
41	6	3.7	241	1	CS31_ECOLI	P15483	escherichia
42	6	3.7	243	1	SDGF_RAT	P24338	rattus norv
43	6	3.7	250	1	VNST_PTPV	P03516	punta toro
44	6	3.7	251	1	YD37_MYCLE	P53426	mycobacteri
45	6	3.7	256	1	DSBG_PSEAE	Q9i106	pseudomonas
46	6	3.7	256	1	FRDB_HAEIN	P44893	haemophilus
47	6	3.7	258	1	RT15_MOUSE	Q9dc71	mus musculu
48	6	3.7	261	1	FLIR_ECOLI	P33135	escherichia
49	6	3.7	265	1	XTMA_BACSU	P39785	bacillus su
50	6	3.7	273	1	OPSG_ODOVI	O18911	odocoileus
51	6	3.7	273	1	OPSR_CANFA	O18914	canis famil
52	6	3.7	273	1	OPSR_HORSE	O18912	equus cabal
53	6	3.7	277	1	XKDB_BACSU	P39781	bacillus su
54	6	3.7	277	1	Y154_ARCFU	O30083	archaeoglob
55	6	3.7	278	1	YFOL_STRTR	P96051	streptococc
56	6	3.7	298	1	PYRD_LACPL	P77887	lactobacill
57	6	3.7	301	1	NHAR_ECOLI	P10087	escherichia
58	6	3.7	301	1	YM77_MYCTU	Q50687	mycobacteri
59	6	3.7	306	1	STX1_CAEEL	Q20024	caenorhabdi
60	6	3.7	308	1	SAL_SILAS	Q9pvw8	silurus aso
61	6	3.7	309	1	FMT_CLOPE	Q8xj13	clostridium
62	6	3.7	314	1	YDGH_ECOLI	P76177	escherichia
63	6	3.7	315	1	MFT_HUMAN	Q9h2d1	homo sapien
64	6	3.7	322	1	MTRH_METKA	O32869	methanopyru
65	6	3.7	330	1	1A1D_PYRAB	Q9v212	pyrococcus
66	6	3.7	333	1	STLK_DROME	P83098	drosophila
67	6	3.7	333	1	Y4MJ_RHISN	P55569	rhizobium s
68	6	3.7	338	1	G3P_TRIHA	P87197	trichoderma
69	6	3.7	338	1	RL3_HALMA	P20279	haloarcula
70	6	3.7	340	1	YDDR_ECOLI	P77308	escherichia
71	6	3.7	341	1	CBFB_RAT	P18576	rattus norv
72	6	3.7	345	1	CATD_PIG	P00795	sus scrofa
73	6	3.7	346	1	CBFB_MOUSE	P23708	mus musculu
74	6	3.7	346	1	YF20_MYCTU	Q50587	mycobacteri
75	6	3.7	347	1	CBFB_HUMAN	P23511	homo sapien
76	6	3.7	350	1	G3P_TREPA	O83816	treponema p
77	6	3.7	350	1	OPSL_CALJA	P34989	callithrix
78	6	3.7	356	1	PROB_AQUAE	O67209	aquifex aeo
79	6	3.7	359	1	PTP4_CAEEL	P34442	caenorhabdi
80	6	3.7	364	1	OPSG_HUMAN	P04001	homo sapien
81	6	3.7	364	1	OPSR_CAPHI	Q95170	capra hircu
82	6	3.7	364	1	OPSR_FELCA	O18913	felis silve
83	6	3.7	364	1	OPSR_HUMAN	P04000	homo sapien
84	6	3.7	371	1	THII_SULSO	Q980g6	sulfolobus
85	6	3.7	376	1	MID2_YEAST	P36027	saccharomyc
86	6	3.7	391	1	CAR1_CANAL	P28872	candida alb
87	6	3.7	391	1	KC21_CHICK	P21868	gallus gall
88	6	3.7	391	1	KC21_HUMAN	P19138	homo sapien
89	6	3.7	391	1	KC21_MOUSE	Q60737	mus musculu
90	6	3.7	391	1	KC21_RABIT	P33674	oryctolagus
91	6	3.7	391	1	KC21_RAT	P19139	rattus norv
92	6	3.7	391	1	YJLD_BACSU	P80861	bacillus su
93	6	3.7	392	1	KC22_XENLA	P28020	xenopus lae
94	6	3.7	395	1	RRPP_PI2H	P23055	human parai
95	6	3.7	395	1	RRPP_PI2HT	P23056	human parai
96	6	3.7	397	1	PGK_ZYMMO	P09404	zymomonas m
97	6	3.7	407	1	RL4A_ARATH	P49691	arabidopsis
98	6	3.7	408	1	ODP2_RICPR	Q9zd20	rickettsia
99	6	3.7	412	1	SDAC_HAEIN	P44615	haemophilus
100	6	3.7	413	1	GAT1_MOUSE	P17679	mus musculu
101	6	3.7	425	1	MNTH_BACSU	P96593	bacillus su
102	6	3.7	426	1	MOEA_MYCTU	O05577	mycobacteri
103	6	3.7	433	1	ANM2_HUMAN	P55345	homo sapien
104	6	3.7	437	1	ACRO_RAT	P29293	rattus norv
105	6	3.7	437	1	GNTT_ECOLI	P39835	escherichia
106	6	3.7	437	1	INTR_SACER	P22877	saccharopol

107	6	3.7	445	1	YG2V_YEAST	P53264	saccharomyc
108	6	3.7	448	1	ANM2_MOUSE	Q9r144	mus musculu
109	6	3.7	450	1	G6PI_CLOPE	Q8xi54	clostridium
110	6	3.7	455	1	HYEP_HUMAN	P07099	homo sapien
111	6	3.7	455	1	PHR_STRGR	P12768	streptomyce
112	6	3.7	456	1	G64A_DROME	P83293	drosophila
113	6	3.7	456	1	NY5R_RAT	Q63634	rattus norv
114	6	3.7	456	1	YC88_MYCTU	Q10614	mycobacteri
115	6	3.7	462	1	PTYC_ECOLI	P32059	escherichia
116	6	3.7	466	1	NY5R_MOUSE	O70342	mus musculu
117	6	3.7	467	1	ATPX_BACFI	P25075	bacillus fi
118	6	3.7	470	1	ATPB_BACFI	P22478	bacillus fi
119	6	3.7	482	1	MTTB_METTE	Q9p995	methanosarc
120	6	3.7	483	1	K1CL_MOUSE	Q64291	mus musculu
121	6	3.7	484	1	UL32_HSV6U	P52463	human herpe
122	6	3.7	492	1	PDI1_SCHPO	Q10057	schizosacch
123	6	3.7	495	1	E1BL_ADE02	P03244	human adeno
124	6	3.7	496	1	E1BL_ADE05	P03243	human adeno
125	6	3.7	499	1	AMPA_CHLTR	O84049	chlamydia t
126	6	3.7	523	1	PUR9_MYCTU	P71553	m bifunctio
127	6	3.7	524	1	CLPP_CHLRE	P42380	chlamydomon
128	6	3.7	527	1	PUR9_MYCLE	Q9z5h5	m bifunctio
129	6	3.7	541	1	PELP_YERPS	P11278	yersinia ps
130	6	3.7	558	1	Y322_MYCGE	P47564	mycoplasma
131	6	3.7	568	1	PELP_ERWCA	P14005	erwinia car
132	6	3.7	576	1	GRK6_HUMAN	P43250	homo sapien
133	6	3.7	579	1	ILVB_SPIPL	P27868	spirulina p
134	6	3.7	582	1	HEMA_MUMP1	P19762	mumps virus
135	6	3.7	582	1	HEMA_MUMPM	P11235	mumps virus
136	6	3.7	582	1	HEMA_MUMPR	P10866	mumps virus
137	6	3.7	582	1	KICH_YEAST	P20485	saccharomyc
138	6	3.7	582	1	YK85_YEAST	P36172	saccharomyc
139	6	3.7	583	1	T2F1_FLAOK	P14870	flavobacter
140	6	3.7	590	1	GRK5_BOVIN	P43249	bos taurus
141	6	3.7	590	1	GRK5_RAT	Q62833	rattus norv
142	6	3.7	593	1	FA12_BOVIN	P98140	bos taurus
143	6	3.7	610	1	SAN1_YEAST	P22470	saccharomyc
144	6	3.7	612	1	PAG3_CAEEL	P34678	caenorhabdi
145	6	3.7	615	1	FA12_HUMAN	P00748	homo sapien
146	6	3.7	619	1	SELB_HAEIN	P43927	haemophilus
147	6	3.7	623	1	PGMP_ARATH	Q9scy0	arabidopsis
148	6	3.7	623	1	RSD1_YEAST	P32368	saccharomyc
149	6	3.7	633	1	ETF1_FOWPV	P21966	fowlpox vir
150	6	3.7	633	1	Y561_HAEIN	P44016	haemophilus
151	6	3.7	636	1	GSH1_HUMAN	P48506	homo sapien
152	6	3.7	637	1	ETF1_VACCA	O93120	vaccinia vi
153	6	3.7	637	1	ETF1_VACCC	P20634	vaccinia vi
154	6	3.7	637	1	ETF1_VACCT	Q9jfa3	vaccinia vi
155	6	3.7	637	1	ETF1_VACCV	P04308	vaccinia vi
156	6	3.7	637	1	ETF1_VARV	P33056	variola vir
157	6	3.7	639	1	AMYG_ASPAK	P23176	aspergillus
158	6	3.7	639	1	AMYG_ASPSH	P22832	aspergillus
159	6	3.7	640	1	AMYG_ASPNG	P04064	aspergillus
160	6	3.7	663	1	Y4WM_RHISN	P55691	rhizobium s
161	6	3.7	664	1	UL47_HSV1F	P08313	herpes simp
162	6	3.7	667	1	EM70_YEAST	P32802	saccharomyc
163	6	3.7	674	1	NUOL_NEIMA	Q9jx92	neisseria m
164	6	3.7	674	1	NUOL_NEIMB	Q9k1b0	neisseria m
165	6	3.7	690	1	EP42_HUMAN	P16452	homo sapien
166	6	3.7	690	1	EP42_MOUSE	P49222	mus musculu
167	6	3.7	693	1	UL47_HSV11	P10231	herpes simp
168	6	3.7	703	1	UL17_HSV11	P10201	herpes simp
169	6	3.7	710	1	ETF2_VACCA	O57220	vaccinia vi
170	6	3.7	710	1	ETF2_VACCC	P20635	vaccinia vi
171	6	3.7	710	1	ETF2_VACCT	Q9jf93	vaccinia vi
172	6	3.7	710	1	ETF2_VACCV	P20636	vaccinia vi
173	6	3.7	710	1	ETF2_VARV	P33806	variola vir
174	6	3.7	711	1	ETF2_MYXVL	Q9q8k4	myxoma viru
175	6	3.7	713	1	ETF2_YABAM	Q9qb84	yaba monkey
176	6	3.7	722	1	Y174_RICPR	P81171	rickettsia
177	6	3.7	732	1	EF2_THEAC	P26752	thermoplasm
178	6	3.7	754	1	PURL_MYCLE	Q50023	mycobacteri
179	6	3.7	754	1	PURL_MYCTU	P54876	mycobacteri
180	6	3.7	763	1	G6PE_RABIT	P56201	oryctolagus
181	6	3.7	763	1	IF39_YEAST	P06103	saccharomyc
182	6	3.7	770	1	DAB2_HUMAN	P98082	homo sapien
183	6	3.7	783	1	TRF2_THEAC	O93654	thermoplasm
184	6	3.7	791	1	G6PE_HUMAN	O95479	homo sapien
185	6	3.7	826	1	CRAA_BACUH	Q9s597	bacillus th
186	6	3.7	834	1	YMZ2_YEAST	Q03177	saccharomyc
187	6	3.7	839	1	PHLD_BOVIN	P80109	bos taurus
188	6	3.7	850	1	STB2_YEAST	P46679	saccharomyc
189	6	3.7	853	1	VMTH_LAMBD	P03736	bacteriopha
190	6	3.7	904	1	YSX7_CAEEL	Q10025	caenorhabdi
191	6	3.7	921	1	T2D3_DROME	P47825	drosophila
192	6	3.7	930	1	PMP8_CHLPN	Q9z393	chlamydia p
193	6	3.7	930	1	SM6C_HUMAN	Q9h3t2	homo sapien
194	6	3.7	967	1	ATS1_RAT	Q9wuq1	rattus norv
195	6	3.7	967	1	PDB1_ARATH	P93733	arabidopsis
196	6	3.7	968	1	ATS1_MOUSE	P97857	mus musculu
197	6	3.7	969	1	MNR2_YEAST	P35724	saccharomyc
198	6	3.7	1010	1	CLPP_CHLEU	P42379	chlamydomon
199	6	3.7	1018	1	CA26_HUMAN	P12110	homo sapien
200	6	3.7	1029	1	CA26_MOUSE	Q02788	mus musculu
201	6	3.7	1034	1	GCSB_FLAPR	P49362	flaveria pr
202	6	3.7	1034	1	GCSP_FLAAN	O49850	flaveria an
203	6	3.7	1034	1	GCSP_FLATR	O49852	flaveria tr
204	6	3.7	1035	1	GCSP_SOLTU	O49954	solanum tub
205	6	3.7	1037	1	ATC3_SCHPO	P22189	schizosacch
206	6	3.7	1057	1	GCSP_PEA	P26969	pisum sativ
207	6	3.7	1060	1	ZFH1_DROME	P28166	drosophila
208	6	3.7	1081	1	PDR6_YEAST	P32767	saccharomyc
209	6	3.7	1097	1	KPC1_CANAL	P43057	candida alb
210	6	3.7	1131	1	MOG1_CAEEL	P34498	caenorhabdi
211	6	3.7	1139	1	YO60_CAEEL	P34609	caenorhabdi
212	6	3.7	1174	1	CIKE_DROME	Q02280	drosophila
213	6	3.7	1221	1	FBL2_MOUSE	P37889	mus musculu
214	6	3.7	1251	1	YQU3_CAEEL	Q09550	caenorhabdi
215	6	3.7	1281	1	TRPS_MOUSE	Q925h1	mus musculu
216	6	3.7	1374	1	M3K5_HUMAN	Q99683	homo sapien
217	6	3.7	1379	1	M3K5_MOUSE	O35099	mus musculu
218	6	3.7	1387	1	PUR4_ARATH	Q9m8d3	arabidopsis
219	6	3.7	1391	1	RPOB_MYCPN	P78013	mycoplasma
220	6	3.7	1400	1	RON_HUMAN	Q04912	homo sapien
221	6	3.7	1402	1	Y197_MOUSE	Q9z0w3	mus musculu
222	6	3.7	1492	1	DPOA_OXYNO	Q94636	oxytricha n
223	6	3.7	1513	1	DPOA_OXYTR	Q27152	oxytricha t
224	6	3.7	1527	1	CA1H_MOUSE	P39061	mus musculu
225	6	3.7	1564	1	MRP2_RABIT	Q28689	oryctolagus
226	6	3.7	1593	1	AT12_HUMAN	P58397	homo sapien
227	6	3.7	1653	1	YFHM_ECOLI	P76578	escherichia
228	6	3.7	1691	1	POLN_HEVME	Q03495	hepatitis e
229	6	3.7	1897	1	PTPF_HUMAN	P10586	homo sapien
230	6	3.7	1984	1	YL_DROME	P98163	drosophila
231	6	3.7	2035	1	Y233_HUMAN	Q92508	homo sapien
232	6	3.7	2136	1	YCF2_MARPO	P09975	marchantia
233	6	3.7	2209	1	Y166_HUMAN	P50748	homo sapien
234	6	3.7	2514	1	POLN_ONNVG	P13886	o'nyong-nyo
235	6	3.7	3056	1	ATM_HUMAN	Q13315	homo sapien
236	6	3.7	3066	1	ATM_MOUSE	Q62388	mus musculu
237	6	3.7	3149	1	TEGU_EBV	P03186	epstein-bar
238	6	3.7	4085	1	RRPA_CVH22	Q05002	human coron
239	6	3.7	4351	1	FAT2_RAT	O88277	rattus norv
240	6	3.7	4644	1	DYHC_MOUSE	Q9jhu4	mus musculu
241	6	3.7	4644	1	DYHC_RAT	P38650	rattus norv
242	6	3.7	4725	1	DYHC_DICDI	P34036	dictyosteli
243	6	3.7	5217	1	HTS1_COCCA	Q01886	cochliobolu
244	5	3.1	15	1	UC29_MAIZE	P80635	zea mays (m
245	5	3.1	23	1	AU41_LITAU	P82397	litoria aur
246	5	3.1	23	1	AU43_LITAU	P82399	litoria aur
247	5	3.1	24	1	PGQ_XENLA	P39080	xenopus lae
248	5	3.1	25	1	NEUU_RANTE	P20056	rana tempor
249	5	3.1	27	1	Y803_ARCFU	O29455	archaeoglob
250	5	3.1	38	1	PSBX_ODOSI	P49509	odontella s
251	5	3.1	40	1	DLDH_SOLTU	P80503	solanum tub
252	5	3.1	49	1	LEC2_AXIPO	P28587	axinella po

253	5	3.1	51	1	RL33_PSEAE	Q9htn9	pseudomonas
254	5	3.1	54	1	RL33_ECOLI	P02436	escherichia
255	5	3.1	55	1	RL33_VIBCH	Q9kvc7	vibrio chol
256	5	3.1	55	1	RL33_YERPE	Q8zjp1	yersinia pe
257	5	3.1	58	1	AXP2_ANTAF	P81548	anthopleura
258	5	3.1	61	1	CSRA_PSEAE	O69078	pseudomonas
259	5	3.1	63	1	CE3F_HYPCU	P50722	hyphantria
260	5	3.1	63	1	CE3G_HYPCU	P50723	hyphantria
261	5	3.1	63	1	PER_DROIM	Q03294	drosophila
262	5	3.1	63	1	RPO9_FOWPV	Q9j5a4	fowlpox vir
263	5	3.1	63	1	RPO9_MYXVL	Q9q8p6	myxoma viru
264	5	3.1	63	1	RPO9_SFVKA	Q9q921	shope fibro
265	5	3.1	63	1	RPO9_VACCV	Q89560	vaccinia vi
266	5	3.1	63	1	RPO9_VARV	Q07047	variola vir
267	5	3.1	63	1	RPO9_YLDV	Q9dhq8	yaba-like d
268	5	3.1	63	1	TAPB_PSEFR	P72192	pseudomonas
269	5	3.1	64	1	CAPA_PSEFR	P72188	pseudomonas
270	5	3.1	65	1	PER_DROMO	Q03295	drosophila
271	5	3.1	66	1	RPON_SULSO	Q980z8	sulfolobus
272	5	3.1	67	1	CSPF_STRCO	P48859	streptomyce
273	5	3.1	67	1	RPOZ_TREPA	O83699	treponema p
274	5	3.1	67	1	VLYS_BPT3	P10307	bacteriopha
275	5	3.1	67	1	VLYS_BPT7	P03802	bacteriopha
276	5	3.1	67	1	Y7K8_BPP22	P14108	bacteriopha
277	5	3.1	69	1	CSPA_RHIME	Q9z3s6	rhizobium m
278	5	3.1	69	1	PSAE_SYNP2	P31969	synechococc
279	5	3.1	69	1	Y842_ARCFU	O29416	archaeoglob
280	5	3.1	75	1	MCHB_ECOLI	Q9rm53	escherichia
281	5	3.1	77	1	MT2_VICFA	Q41657	vicia faba
282	5	3.1	79	1	MT2_CICAR	Q39459	cicer ariet
283	5	3.1	79	1	PP28_MCMVK	P52368	murine cyto
284	5	3.1	79	1	RL22_CLOPP	Q46229	clover prol
285	5	3.1	79	1	RL22_PRUAP	Q44161	prunus arme
286	5	3.1	80	1	RUXG_ARATH	O82221	arabidopsis
287	5	3.1	80	1	YC15_OENVI	P31566	oenothera v
288	5	3.1	81	1	RUXG_MEDSA	P24715	medicago sa
289	5	3.1	82	1	S61B_ARATH	P38389	arabidopsis
290	5	3.1	83	1	RS5_ARATH	P51427	arabidopsis
291	5	3.1	83	1	YF24_ARCFU	O28748	archaeoglob
292	5	3.1	85	1	RNFH_RHOCA	P97056	rhodobacter
293	5	3.1	86	1	Y572_METJA	Q57992	methanococc
294	5	3.1	87	1	YSCF_YEREN	Q01247	yersinia en
295	5	3.1	88	1	VE4_HPV18	P06791	human papil
296	5	3.1	88	1	WIRA_WHEAT	Q01482	triticum ae
297	5	3.1	90	1	HYPC_ALCEU	P31900	alcaligenes
298	5	3.1	90	1	YT65_CAEEL	Q11081	caenorhabdi
299	5	3.1	91	1	BXB2_SAMCY	P33722	samia cynth
300	5	3.1	91	1	EML_PICGL	Q40864	picea glauc
301	5	3.1	91	1	RL22_ASHYP	Q44593	ash yellows
302	5	3.1	91	1	RL22_LOWBP	Q48879	loofah witc
303	5	3.1	92	1	COPR_STRAG	P24716	streptococc
304	5	3.1	92	1	PLM_HUMAN	O00168	homo sapien
305	5	3.1	93	1	VSMP_IBVP3	P30246	avian infec
306	5	3.1	94	1	RL22_TOBBP	Q56252	tomato big
307	5	3.1	95	1	CH10_CLOAB	P30719	clostridium
308	5	3.1	95	1	CH10_FRATU	P94797	francisella
309	5	3.1	95	1	RL34_AERPE	P58026	aeropyrum p
310	5	3.1	97	1	NB2M_HUMAN	O43676	homo sapien
311	5	3.1	97	1	YDB8_SCHPO	Q10360	schizosacch
312	5	3.1	98	1	S111_MOUSE	P50543	mus musculu
313	5	3.1	99	1	CH10_MYCAV	O86017	mycobacteri
314	5	3.1	99	1	CH10_MYCBO	P15020	mycobacteri
315	5	3.1	99	1	CH10_MYCLE	P24301	mycobacteri
316	5	3.1	99	1	CH10_MYCTU	P09621	mycobacteri
317	5	3.1	99	1	PD11_BOVIN	P56982	bos taurus
318	5	3.1	99	1	YM23_MARPO	P38466	marchantia
319	5	3.1	100	1	RL23_ACTAC	P55839	actinobacil
320	5	3.1	100	1	URE3_HAEIN	P44393	haemophilus
321	5	3.1	100	1	URE3_STRSL	Q55053	streptococc
322	5	3.1	100	1	Y908_HAEIN	P44073	haemophilus
323	5	3.1	101	1	GLR1_SCHPO	O36032	schizosacch
324	5	3.1	101	1	PLF4_HUMAN	P02776	homo sapien
325	5	3.1	101	1	S107_BOVIN	Q28050	bos taurus
326	5	3.1	101	1	VG30_BPML5	Q05239	mycobacteri
327	5	3.1	102	1	CH10_STRAL	Q00769	streptomyce
328	5	3.1	102	1	CH10_STRCO	P40172	streptomyce
329	5	3.1	102	1	S111_RABIT	P24480	oryctolagus
330	5	3.1	102	1	S114_HUMAN	O60417	homo sapien
331	5	3.1	103	1	ES6D_MYCTU	O05440	mycobacteri
332	5	3.1	103	1	HIS2_RHOSH	P50935	rhodobacter
333	5	3.1	103	1	KAC4_RABIT	P01840	oryctolagus
334	5	3.1	103	1	RL24_HAEIN	P44362	haemophilus
335	5	3.1	103	1	VC20_VACCC	P21104	vaccinia vi
336	5	3.1	103	1	YG1R_YEAST	P53225	saccharomyc
337	5	3.1	104	1	PF4V_HUMAN	P10720	homo sapien
338	5	3.1	104	1	S11Y_HUMAN	Q9udp3	homo sapien
339	5	3.1	105	1	PLF4_RAT	P06765	rattus norv
340	5	3.1	105	1	RL21_RICPR	Q9zci9	rickettsia
341	5	3.1	105	1	S111_HUMAN	P31949	homo sapien
342	5	3.1	105	1	Y756_RICPR	Q9zci4	rickettsia
343	5	3.1	105	1	YM74_MYCTU	Q50689	mycobacteri
344	5	3.1	106	1	VSMP_IBVU4	P30247	avian infec
345	5	3.1	106	1	Y547_RICPR	Q9zd03	rickettsia
346	5	3.1	107	1	GRLA_RICPR	O05957	rickettsia
347	5	3.1	107	1	ULB1_HCMVA	P16831	human cytom
348	5	3.1	107	1	VSMP_IBVU5	P30248	avian infec
349	5	3.1	107	1	YB11_BACHD	Q9kdv0	bacillus ha
350	5	3.1	107	1	YJ09_BRUME	Q8yeg9	brucella me
351	5	3.1	108	1	KRFJ_CHICK	O13152	gallus gall
352	5	3.1	108	1	RSN_HUMAN	Q9hd89	h resistin
353	5	3.1	108	1	YNFA_ECO57	Q8x7a6	escherichia
354	5	3.1	108	1	YNFA_ECOLI	P76169	escherichia
355	5	3.1	108	1	YNFA_SALTI	Q8z6z2	salmonella
356	5	3.1	108	1	YNFA_SALTY	Q8zpj2	salmonella
357	5	3.1	108	1	YR44_BACHD	Q9k9a5	bacillus ha
358	5	3.1	109	1	VSMP_IBVK	P19744	avian infec
359	5	3.1	109	1	Y268_CAUCR	Q9abf9	caulobacter
360	5	3.1	109	1	YFJF_BACSU	O31553	bacillus su
361	5	3.1	110	1	RLA2_SCHPO	P08094	schizosacch
362	5	3.1	110	1	RLA4_YEAST	P02400	saccharomyc
363	5	3.1	110	1	Y332_METJA	Q57778	methanococc
364	5	3.1	110	1	YC27_ARCFU	O29041	archaeoglob
365	5	3.1	110	1	YQ39_MYCTU	P71938	mycobacteri
366	5	3.1	110	1	YZ25_METJA	Q60284	methanococc
367	5	3.1	111	1	IBP5_BOVIN	Q05717	bos taurus
368	5	3.1	111	1	LECS_RANJA	P18839	rana japoni
369	5	3.1	111	1	LY6G_MOUSE	P35461	mus musculu
370	5	3.1	111	1	RL22_ACHLA	P29222	acholeplasm
371	5	3.1	111	1	YG2C_YEAST	P53245	saccharomyc
372	5	3.1	112	1	ANFG_AZOMA	O68946	azomonas ma
373	5	3.1	112	1	GLNB_RHOSH	P43519	rhodobacter
374	5	3.1	112	1	LSR2_MYCLE	P24094	mycobacteri
375	5	3.1	112	1	LSR2_MYCTU	O06285	mycobacteri
376	5	3.1	112	1	PEDB_PEDAC	P36496	pediococcus
377	5	3.1	113	1	RL22_THETH	P48286	thermus the
378	5	3.1	113	1	SSIF_STRBI	Q9r645	streptomyce
379	5	3.1	113	1	V186_FOWPV	Q9j548	fowlpox vir
380	5	3.1	113	1	YC06_MYCPN	P75570	mycoplasma
381	5	3.1	114	1	CU05_HOMAM	P81389	homarus ame
382	5	3.1	116	1	ANFG_CLOPA	Q46244	clostridium
383	5	3.1	116	1	AQN3_PIG	P24020	sus scrofa
384	5	3.1	116	1	NU3M_CROLA	P34192	crossostoma
385	5	3.1	116	1	RL20_HAEIN	P44358	haemophilus
386	5	3.1	116	1	YZ06_METJA	Q60261	methanococc
387	5	3.1	117	1	NLT5_ORYSA	O65091	oryza sativ
388	5	3.1	117	1	RL20_ECOLI	P02421	escherichia
389	5	3.1	117	1	RL20_PASMU	Q9cn41	pasteurella
390	5	3.1	117	1	RL20_VIBCH	O68846	vibrio chol
391	5	3.1	117	1	Y12K_RHDV	P27412	rabbit hemo
392	5	3.1	117	1	Y12K_RHDV3	P27413	rabbit hemo
393	5	3.1	118	1	R18E_SULTO	Q96yw1	sulfolobus
394	5	3.1	118	1	RL20_BUCAI	P57228	buchnera ap
395	5	3.1	118	1	RL20_PSEAE	Q9i0a2	pseudomonas
396	5	3.1	118	1	RL20_PSEFL	Q9x6e8	pseudomonas
397	5	3.1	118	1	RL20_PSESY	P52828	pseudomonas
398	5	3.1	118	1	RL20_YERPE	Q8zdw8	yersinia pe

399	5	3.1	118	1	SYC_STAXY	P77986	staphylococ
400	5	3.1	119	1	AMCY_METEX	P04172	methylobact
401	5	3.1	119	1	CHA1_ANTPO	P02846	antheraea p
402	5	3.1	119	1	FOLB_STRCO	Q9x8i0	streptomyce
403	5	3.1	119	1	NQO7_THETH	Q56217	thermus the
404	5	3.1	119	1	RNP_IGUIG	P80287	iguana igua
405	5	3.1	119	1	WN1A_PLEJO	P28131	plethodon j
406	5	3.1	120	1	C550_BACSU	P24469	bacillus su
407	5	3.1	120	1	WN1A_ALOVU	P28099	alopias vul
408	5	3.1	121	1	GSPI_KLEPN	P15748	klebsiella
409	5	3.1	122	1	YBDF_ECOLI	P39454	escherichia
410	5	3.1	123	1	YQ30_MYCLE	O06091	mycobacteri
411	5	3.1	124	1	MPP5_HUMAN	Q99546	homo sapien
412	5	3.1	124	1	NEUR_FELCA	P41737	felis silve
413	5	3.1	124	1	PER_HIRPI	Q25109	hirtodrosop
414	5	3.1	124	1	RL14_MYCGA	O52342	mycoplasma
415	5	3.1	124	1	YKF3_YEAST	P35733	saccharomyc
416	5	3.1	125	1	GSPI_ECOLI	P45760	escherichia
417	5	3.1	126	1	YQHL_BACSU	P54510	bacillus su
418	5	3.1	129	1	NUSB_STAAM	Q99tw9	staphylococ
419	5	3.1	129	1	RL22_ASTYP	Q50266	aster yello
420	5	3.1	129	1	RL22_PHYS2	O66097	phytoplasma
421	5	3.1	130	1	ANFG_CLOHU	Q46083	clostridium
422	5	3.1	130	1	COTM_BACSU	Q45058	bacillus su
423	5	3.1	130	1	CRCB_SYNY3	P72836	synechocyst
424	5	3.1	130	1	INI2_HUMAN	P09912	homo sapien
425	5	3.1	130	1	INI2_PANTR	Q28808	pan troglod
426	5	3.1	130	1	LEG2_MOUSE	Q9cqw5	mus musculu
427	5	3.1	130	1	LEG2_RAT	Q9z144	rattus norv
428	5	3.1	130	1	OSP4_PEA	P17771	pisum sativ
429	5	3.1	131	1	NUSB_BACSU	P54520	bacillus su
430	5	3.1	131	1	PER_ZAPTU	Q27135	zaprionus t
431	5	3.1	131	1	SP27_CAEEL	P54218	caenorhabdi
432	5	3.1	131	1	YQHP_BACSU	P54514	bacillus su
433	5	3.1	132	1	ANFG_AZOPA	O68943	azotobacter
434	5	3.1	132	1	ANFG_AZOVI	P16268	azotobacter
435	5	3.1	132	1	SVP2_CAVPO	P16531	cavia porce
436	5	3.1	133	1	AMPE_RAT	P50123	rattus norv
437	5	3.1	133	1	HEMI_PYRAP	P37363	pyrrhocoris
438	5	3.1	133	1	Y044_BORBU	O51073	borrelia bu
439	5	3.1	134	1	LY6I_MOUSE	Q9wu67	mus musculu
440	5	3.1	134	1	YHI1_MYCLE	P49774	mycobacteri
441	5	3.1	134	1	YM86_YEAST	Q04814	saccharomyc
442	5	3.1	135	1	Y4PE_RHISN	P55614	rhizobium s
443	5	3.1	136	1	BLS_STRSJ	P19997	streptovert
444	5	3.1	136	1	RS6E_METAC	Q8tql4	methanosarc
445	5	3.1	136	1	VG35_BPMU	Q9t1w0	bacteriopha
446	5	3.1	136	1	YIO1_CVBM	P22053	bovine coro
447	5	3.1	137	1	ATPE_SYNP7	P07892	synechococc
448	5	3.1	137	1	MCRD_METTM	P11564	methanobact
449	5	3.1	137	1	RL32_TRIHA	P34040	trichoderma
450	5	3.1	137	1	SPMI_PIG	Q28920	sus scrofa
451	5	3.1	138	1	Y29A_MYCGE	Q9zb76	mycoplasma
452	5	3.1	138	1	YFFO_ECOLI	P76546	escherichia
453	5	3.1	138	1	YMSP_IRV22	P25097	simulium ir
454	5	3.1	140	1	VMTG_LAMBD	P03734	bacteriopha
455	5	3.1	141	1	DUT_CHVP1	O41033	paramecium
456	5	3.1	141	1	RK11_CYAPA	P48126	cyanophora
457	5	3.1	141	1	TATB_PSEAE	Q9hub4	pseudomonas
458	5	3.1	141	1	VTU1_DROME	P11449	drosophila
459	5	3.1	141	1	X_WHV59	P12914	woodchuck h
460	5	3.1	141	1	X_WHV7	P12913	woodchuck h
461	5	3.1	141	1	X_WHV8	P12937	woodchuck h
462	5	3.1	141	1	X_WHV8I	P17401	woodchuck h
463	5	3.1	141	1	X_WHVW6	P11294	woodchuck h
464	5	3.1	141	1	YPEA_ECOLI	P76539	escherichia
465	5	3.1	142	1	LYSX_DROME	P37161	drosophila
466	5	3.1	142	1	NXT2_HUMAN	Q9npj8	homo sapien
467	5	3.1	143	1	YEN6_YEAST	P40031	saccharomyc
468	5	3.1	144	1	NU6M_ASCSU	P24873	ascaris suu
469	5	3.1	144	1	R12A_ARATH	Q9s9p1	arabidopsis
470	5	3.1	144	1	R12C_ARATH	Q9skz3	arabidopsis
471	5	3.1	145	1	DTD_DEIRA	Q9rvy1	deinococcus
472	5	3.1	145	1	DUT_CLOAB	Q97j61	clostridium
473	5	3.1	145	1	LIRP_LOCMI	P15131	locusta mig
474	5	3.1	145	1	MMF1_YEAST	P40185	saccharomyc
475	5	3.1	145	1	Y883_RICPR	Q9zc85	rickettsia
476	5	3.1	145	1	YAF4_SCHPO	Q09858	schizosacch
477	5	3.1	145	1	YCEA_PAELA	P29718	paenibacill
478	5	3.1	146	1	E316_ADE03	P11320	human adeno
479	5	3.1	146	1	RBFA_RHILO	Q98bi7	rhizobium l
480	5	3.1	146	1	RL15_MYCLE	O33002	mycobacteri
481	5	3.1	147	1	RCP1_SYNY3	Q55169	synechocyst
482	5	3.1	147	1	RL15_PYRAB	Q9v1v7	pyrococcus
483	5	3.1	147	1	RL15_SYNY3	P73303	synechocyst
484	5	3.1	148	1	MOAE_STAAM	Q99s01	staphylococ
485	5	3.1	148	1	YPOL_IPNVJ	P22931	infectious
486	5	3.1	149	1	ARGR_BACHD	Q9k973	bacillus ha
487	5	3.1	149	1	YB29_MYCPN	P75346	mycoplasma
488	5	3.1	150	1	MOAE_STACA	Q9zim9	staphylococ
489	5	3.1	150	1	YEAL_HAEIN	P44110	haemophilus
490	5	3.1	150	1	YP59_YEAST	Q02784	saccharomyc
491	5	3.1	151	1	DRPD_CRAPL	P22241	craterostig
492	5	3.1	151	1	MER3_EUPOC	P28717	euplotes oc
493	5	3.1	151	1	OM22_YEAST	P49334	saccharomyc
494	5	3.1	152	1	RK34_SPIOL	P82244	spinacia ol
495	5	3.1	153	1	FLAG_VIBCH	Q9kq62	vibrio chol
496	5	3.1	153	1	MAL_CANFA	Q28296	canis famil
497	5	3.1	153	1	RL15_TREPA	O83237	treponema p
498	5	3.1	153	1	Y264_METJA	Q57712	methanococc
499	5	3.1	153	1	Y992_SYNY3	P74516	synechocyst
500	5	3.1	153	1	YJEE_ECOLI	P31805	escherichia
501	5	3.1	154	1	DAU1_DAUCA	O04298	daucus caro
502	5	3.1	154	1	HS23_MAIZE	Q08275	zea mays (m
503	5	3.1	154	1	YL66_ARCFU	O28116	archaeoglob
504	5	3.1	155	1	RS7_UREPA	Q9ppw6	ureaplasma
505	5	3.1	155	1	RS9_RHIME	Q92qr5	rhizobium m
506	5	3.1	155	1	Y875_RHILO	Q98lu3	rhizobium l
507	5	3.1	156	1	BCCP_ECOLI	P02905	escherichia
508	5	3.1	156	1	LUXS_STAAM	Q99sc5	staphylococ
509	5	3.1	156	1	NUDH_CAMJE	Q9pht5	campylobact
510	5	3.1	156	1	RS7_ANASP	Q8yp61	anabaena sp
511	5	3.1	157	1	PA1F_HUMAN	P24666	homo sapien
512	5	3.1	157	1	RL12_CHLRE	P50884	chlamydomon
513	5	3.1	157	1	SMP1_HUMAN	O95807	homo sapien
514	5	3.1	157	1	YD74_BACHD	Q9kd45	bacillus ha
515	5	3.1	158	1	CU14_HUMAN	P56557	homo sapien
516	5	3.1	159	1	ALL2_APIGR	P92918	apium grave
517	5	3.1	159	1	GRPA_MEDFA	Q09134	medicago fa
518	5	3.1	159	1	NUDD_ECO57	Q8x7n1	escherichia
519	5	3.1	159	1	NUDD_ECOLI	P32056	escherichia
520	5	3.1	159	1	PPAC_RAT	P41498	rattus norv
521	5	3.1	160	1	FMAJ_BACNO	P19528	bacteroides
522	5	3.1	161	1	ATPF_BUCAI	P57120	buchnera ap
523	5	3.1	161	1	T2DA_HUMAN	Q16514	homo sapien
524	5	3.1	161	1	Y255_BUCAI	P57343	buchnera ap
525	5	3.1	162	1	GCSH_ORYSA	O22535	oryza sativ
526	5	3.1	162	1	LACC_FELCA	P33688	felis silve
527	5	3.1	162	1	Y480_TREPA	O83493	treponema p
528	5	3.1	162	1	YS11_DROME	Q9w0y2	drosophila
529	5	3.1	163	1	DNE1_CHLRE	P05725	chlamydomon
530	5	3.1	163	1	DUT_ADEG8	Q9yys0	avian adeno
531	5	3.1	165	1	R10A_PIG	P53027	sus scrofa
532	5	3.1	165	1	RL12_SCHPO	O75000	schizosacch
533	5	3.1	166	1	DUSP_FOWPV	Q9j592	fowlpox vir
534	5	3.1	166	1	RL12_ARATH	P50883	arabidopsis
535	5	3.1	166	1	RL12_PRUAR	O50003	prunus arme
536	5	3.1	167	1	ET3_RAT	P13207	rattus norv
537	5	3.1	167	1	NU6M_POLOR	Q95919	polypterus
538	5	3.1	168	1	DH1_MAIZE	P12950	zea mays (m
539	5	3.1	168	1	SSB_HAEIN	P44409	haemophilus
540	5	3.1	169	1	ILVH_GUITH	O78451	guillardia
541	5	3.1	169	1	MINC_DEIRA	Q9rtk7	deinococcus
542	5	3.1	169	1	VAL1_CAEEL	Q21898	caenorhabdi
543	5	3.1	169	1	YFJ6_YEAST	P43604	saccharomyc
544	5	3.1	170	1	LE23_METJA	Q58667	methanococc

545	5	3.1	170	1	MIR3_MOUSE	Q9wtw3	mus musculu
546	5	3.1	170	1	OBPA_HUMAN	Q9ny56	homo sapien
547	5	3.1	170	1	OBPB_HUMAN	Q9nph6	homo sapien
548	5	3.1	171	1	AR20_YEAST	P33204	saccharomyc
549	5	3.1	171	1	YB74_YEAST	P38320	saccharomyc
550	5	3.1	171	1	YTUX_AGRVI	P70790	agrobacteri
551	5	3.1	172	1	ILVH_CYACA	Q9tly1	cyanidium c
552	5	3.1	172	1	ILVH_METJA	Q57625	methanococc
553	5	3.1	172	1	MRED_BACSU	Q01467	bacillus su
554	5	3.1	172	1	YHBO_ECOLI	P45470	escherichia
555	5	3.1	173	1	CD3D_MOUSE	P04235	mus musculu
556	5	3.1	173	1	CHB2_LYMDI	P50604	lymantria d
557	5	3.1	173	1	CRGC_HUMAN	P07315	homo sapien
558	5	3.1	173	1	CRGC_MOUSE	Q61597	mus musculu
559	5	3.1	173	1	CRGC_RAT	P02529	rattus norv
560	5	3.1	173	1	CRGE_BOVIN	Q28088	bos taurus
561	5	3.1	173	1	Y096_NPVAC	P41656	autographa
562	5	3.1	173	1	YCED_ECOLI	P14189	escherichia
563	5	3.1	174	1	CRGA_BOVIN	P02527	bos taurus
564	5	3.1	175	1	8513_TRYCR	P18271	trypanosoma
565	5	3.1	175	1	SSB_SERMA	P25762	serratia ma
566	5	3.1	176	1	CON8_NEUCR	P10169	neurospora
567	5	3.1	176	1	PYRE_METJA	Q58509	methanococc
568	5	3.1	176	1	RNMG_ASPRE	P04389	aspergillus
569	5	3.1	176	1	VG03_BPT4	P13331	bacteriopha
570	5	3.1	176	1	Y742_SYNY3	P72687	synechocyst
571	5	3.1	177	1	RNAS_ASPGI	P00655	aspergillus
572	5	3.1	177	1	RNCL_ASPCL	P49074	aspergillus
573	5	3.1	177	1	SSB_ECOLI	P02339	escherichia
574	5	3.1	178	1	CASK_RAT	P04468	rattus norv
575	5	3.1	178	1	DUT_ADEG1	Q89662	avian adeno
576	5	3.1	178	1	NUIC_ORYSA	P12099	oryza sativ
577	5	3.1	179	1	ATPD_THIFE	P41170	thiobacillu
578	5	3.1	179	1	DYR_MYCPU	Q98q32	mycoplasma
579	5	3.1	179	1	Y459_HAEIN	P44722	haemophilus
580	5	3.1	179	1	YAJI_ECOLI	P46122	escherichia
581	5	3.1	179	1	YCBQ_ECOLI	P75855	escherichia
582	5	3.1	179	1	YKHA_BACSU	P49851	bacillus su
583	5	3.1	179	1	YPOL_THETH	P32438	thermus the
584	5	3.1	180	1	NUIC_MAIZE	P46722	zea mays (m
585	5	3.1	180	1	NUIC_WHEAT	P05312	triticum ae
586	5	3.1	180	1	SP22_CAEEL	P34525	caenorhabdi
587	5	3.1	180	1	Y271_METJA	Q57719	methanococc
588	5	3.1	181	1	ARL1_DROME	P25160	drosophila
589	5	3.1	181	1	HB2T_HUMAN	P04232	homo sapien
590	5	3.1	181	1	ILVH_GALSU	Q9ms98	galdieria s
591	5	3.1	181	1	OM20_NEUCR	P35848	neurospora
592	5	3.1	181	1	VM22_YEAST	P38784	saccharomyc
593	5	3.1	181	1	YGGT_HAEIN	P44097	haemophilus
594	5	3.1	182	1	CD3G_HUMAN	P09693	homo sapien
595	5	3.1	182	1	MCRC_METVO	P11567	methanococc
596	5	3.1	182	1	RIMM_ECOLI	P21504	escherichia
597	5	3.1	182	1	RIMM_SALTY	P36246	salmonella
598	5	3.1	182	1	YN83_SULTO	Q96xy5	sulfolobus
599	5	3.1	183	1	HPRT_LACLA	Q02522	lactococcus
600	5	3.1	184	1	ATPF_MARPO	P06291	marchantia
601	5	3.1	184	1	ATPF_PINTH	O62939	pinus thunb
602	5	3.1	184	1	DHLE_RAPSA	P21298	raphanus sa
603	5	3.1	184	1	IPYR_MYCPN	P75250	mycoplasma
604	5	3.1	185	1	DP1_HUMAN	Q00765	homo sapien
605	5	3.1	185	1	DP1_MOUSE	Q60870	mus musculu
606	5	3.1	185	1	VATE_THEAC	Q9hm68	thermoplasm
607	5	3.1	186	1	EFP_STRPN	Q97se8	streptococc
608	5	3.1	186	1	NOS1_SHEEP	Q29498	ovis aries
609	5	3.1	186	1	PTH_UREPA	Q9pr67	ureaplasma
610	5	3.1	187	1	ABP1_TOBAC	P33490	nicotiana t
611	5	3.1	187	1	ABP2_TOBAC	P33491	nicotiana t
612	5	3.1	187	1	KORC_ARCFU	O29779	archaeoglob
613	5	3.1	187	1	SURE_TREPA	O83434	treponema p
614	5	3.1	188	1	COAT_PHMV	P36351	physalis mo
615	5	3.1	188	1	RECX_MYCSM	P94965	mycobacteri
616	5	3.1	188	1	SSX2_HUMAN	Q16385	homo sapien
617	5	3.1	188	1	SSX3_HUMAN	Q99909	homo sapien
618	5	3.1	189	1	APOD_MOUSE	P51910	mus musculu
619	5	3.1	189	1	APOD_RAT	P23593	rattus norv
620	5	3.1	189	1	COAT_CCMV	P03601	cowpea chlo
621	5	3.1	189	1	COPZ_YEAST	P53600	saccharomyc
622	5	3.1	190	1	C24A_BISBI	Q95173	b cytochrom
623	5	3.1	190	1	C24A_BOVIN	O46521	bos taurus
624	5	3.1	190	1	COAT_BMDV	P15158	belladonna
625	5	3.1	190	1	SOMA_LAMPA	P37885	lama guanic
626	5	3.1	190	1	Y516_AERPE	Q9yer5	aeropyrum p
627	5	3.1	190	1	YCX6_YEAST	P25651	saccharomyc
628	5	3.1	191	1	CHB1_LYMDI	P43515	lymantria d
629	5	3.1	191	1	COXZ_RHOSH	P56940	rhodobacter
630	5	3.1	191	1	EAR_EBV	P03182	epstein-bar
631	5	3.1	191	1	Y611_METJA	Q58028	methanococc
632	5	3.1	191	1	YGCP_ECOLI	Q46906	escherichia
633	5	3.1	191	1	YRAP_ECOLI	P45467	escherichia
634	5	3.1	192	1	DENR_CAEEL	Q9nah4	caenorhabdi
635	5	3.1	192	1	GCH2_HELPJ	Q9zl42	helicobacte
636	5	3.1	192	1	GCH2_HELPY	O08315	helicobacte
637	5	3.1	192	1	PABA_STRLI	P27627	streptomyce
638	5	3.1	192	1	RS7_ANOGA	P33514	anopheles g
639	5	3.1	194	1	HIS7_BACHD	Q9k6z3	bacillus ha
640	5	3.1	194	1	RL9_RHILO	Q984u0	rhizobium l
641	5	3.1	195	1	RNFB_VIBCH	Q9kt87	vibrio chol
642	5	3.1	196	1	DEF_LACLA	Q48661	lactococcus
643	5	3.1	196	1	NODA_RHILV	P04338	rhizobium l
644	5	3.1	196	1	RNH2_AQUAE	O67768	aquifex aeo
645	5	3.1	196	1	VATL_DICDI	P54642	dictyosteli
646	5	3.1	196	1	YS64_CAEEL	Q09380	caenorhabdi
647	5	3.1	197	1	CLP3_ANASP	Q8yp43	anabaena sp
648	5	3.1	197	1	DCD_METTH	O27875	methanobact
649	5	3.1	197	1	FCPC_PHATR	Q08586	phaeodactyl
650	5	3.1	197	1	FCPD_PHATR	Q08587	phaeodactyl
651	5	3.1	197	1	PAD1_ECO57	Q9x728	escherichia
652	5	3.1	197	1	UL56_HSV11	P10240	herpes simp
653	5	3.1	198	1	COX2_NOTPE	O03892	nothoprocta
654	5	3.1	198	1	HIS5_METTH	O27568	methanobact
655	5	3.1	198	1	MCRC_METVA	P07960	methanococc
656	5	3.1	198	1	RK27_ARATH	Q9fln4	arabidopsis
657	5	3.1	198	1	XA1_XENLA	P23507	xenopus lae
658	5	3.1	199	1	PEXH_YEAST	P40155	saccharomyc
659	5	3.1	200	1	MCRV_METJA	Q58254	methanococc
660	5	3.1	201	1	DRPF_CRAPL	P23283	craterostig
661	5	3.1	201	1	LASI_PSEAE	P33883	pseudomonas
662	5	3.1	201	1	MGF_CHICK	P13854	gallus gall
663	5	3.1	201	1	NING_BPH19	O48427	bacteriopha
664	5	3.1	201	1	TRPG_VIBCH	Q9kst3	vibrio chol
665	5	3.1	202	1	ADHR_DROER	P28485	drosophila
666	5	3.1	202	1	HIS7_CRYNE	P40919	cryptococcu
667	5	3.1	202	1	VIC2_AGRTU	P06666	agrobacteri
668	5	3.1	202	1	Y740_ARCFU	O29518	archaeoglob
669	5	3.1	203	1	3MGH_MYCTU	O33190	mycobacteri
670	5	3.1	203	1	IF3_HELPJ	Q9zmv2	helicobacte
671	5	3.1	203	1	IF3_HELPY	P55973	helicobacte
672	5	3.1	203	1	YC11_AQUAE	O67264	aquifex aeo
673	5	3.1	204	1	CAHB_MOUSE	O70354	mus musculu
674	5	3.1	204	1	SRPP_HEVBR	O82803	hevea brasi
675	5	3.1	204	1	YC42_ODOSI	P49537	odontella s
676	5	3.1	205	1	HFD2_HAEIN	P45993	haemophilus
677	5	3.1	205	1	RUVA_MYCGE	Q49424	mycoplasma
678	5	3.1	205	1	TRPF_THEMA	Q56320	thermotoga
679	5	3.1	205	1	VA5_VESSQ	P35786	vespula squ
680	5	3.1	205	1	Y054_AQUAE	O66466	aquifex aeo
681	5	3.1	206	1	RL3_THETH	P52860	thermus the
682	5	3.1	206	1	YGZC_YEAST	P53058	saccharomyc
683	5	3.1	207	1	FGFG_HUMAN	O43320	homo sapien
684	5	3.1	207	1	FGFG_RAT	O54769	rattus norv
685	5	3.1	207	1	RASL_CAEEL	P34443	caenorhabdi
686	5	3.1	207	1	RL3_THEMA	P38515	thermotoga
687	5	3.1	207	1	Y214_MYCGE	P47456	mycoplasma
688	5	3.1	208	1	HIS7_TRIHA	P34041	trichoderma
689	5	3.1	208	1	PSB5_HUMAN	P28074	h proteasom
690	5	3.1	208	1	XLR_MOUSE	P05531	mus musculu

691	5	3.1	208	1	Y146_UREPA	Q9pqz8	ureaplasma
692	5	3.1	209	1	ADPP_ECOLI	P36651	escherichia
693	5	3.1	209	1	CLD4_CERAE	O19005	cercopithec
694	5	3.1	209	1	CLD4_HUMAN	O14493	homo sapien
695	5	3.1	209	1	IDI_AGRRH	Q9kwd1	agrobacteri
696	5	3.1	209	1	NQRD_VIBAL	Q57095	vibrio algi
697	5	3.1	209	1	VS10_ROTBS	P34718	bovine rota
698	5	3.1	209	1	YC29_PORPU	P51343	porphyra pu
699	5	3.1	210	1	ACUA_BACSU	P39065	bacillus su
700	5	3.1	210	1	ALKH_TREPA	O83578	t putative
701	5	3.1	210	1	CYC4_PSEST	Q52369	pseudomonas
702	5	3.1	210	1	MOBC_BPT4	P07074	bacteriopha
703	5	3.1	210	1	NQRD_VIBHA	Q9rfv8	vibrio harv
704	5	3.1	210	1	YADS_AERPU	Q9r9s0	aeromonas p
705	5	3.1	210	1	YI50_AQUAE	O67702	aquifex aeo
706	5	3.1	211	1	CES2_CAEEL	Q94126	caenorhabdi
707	5	3.1	211	1	END3_ECOLI	P20625	escherichia
708	5	3.1	211	1	T402_BURCE	P24536	burkholderi
709	5	3.1	211	1	Y359_TREPA	O83378	treponema p
710	5	3.1	211	1	YFB0_YEAST	P43582	saccharomyc
711	5	3.1	211	1	YKGB_PASHA	P46928	pasteurella
712	5	3.1	212	1	PUR3_HAEIN	P43846	haemophilus
713	5	3.1	212	1	Y027_BORBU	O51058	borrelia bu
714	5	3.1	212	1	YG29_HAEIN	P45280	haemophilus
715	5	3.1	213	1	CG1B_MEDSA	P30286	medicago sa
716	5	3.1	213	1	CLP1_PSEAE	Q9i2u1	pseudomonas
717	5	3.1	213	1	CPCF_SYNEL	P50038	synechococc
718	5	3.1	213	1	HS26_YEAST	P15992	saccharomyc
719	5	3.1	213	1	ILL1_HUMAN	P15814	homo sapien
720	5	3.1	213	1	KTHY_ECOLI	P37345	escherichia
721	5	3.1	213	1	RB4B_HUMAN	P22750	homo sapien
722	5	3.1	213	1	RB4B_RAT	P51146	rattus norv
723	5	3.1	214	1	3MGH_MYCLE	O05678	mycobacteri
724	5	3.1	214	1	ET3_MOUSE	P48299	mus musculu
725	5	3.1	214	1	GAL1_CANMA	P56599	candida mal
726	5	3.1	214	1	IGF2_ONCMY	Q02816	oncorhynchu
727	5	3.1	214	1	RB11_DICDI	P36412	dictyosteli
728	5	3.1	214	1	VM24_BORHE	P32778	borrelia he
729	5	3.1	215	1	ENGB_PSEAE	Q9ht81	pseudomonas
730	5	3.1	215	1	RIBB_BUCAI	P57167	buchnera ap
731	5	3.1	215	1	VM03_BORHE	Q02448	borrelia he
732	5	3.1	215	1	VM12_YEAST	P32341	saccharomyc
733	5	3.1	216	1	ATKC_HALN1	P57687	halobacteri
734	5	3.1	216	1	HFD1_HAEIN	P45992	haemophilus
735	5	3.1	216	1	HIS7_SCHPO	P40374	schizosacch
736	5	3.1	216	1	R10A_ICTPU	Q90yv8	ictalurus p
737	5	3.1	216	1	VA36_VARV	P33852	variola vir
738	5	3.1	217	1	OPBB_BACSU	Q45461	bacillus su
739	5	3.1	217	1	OPCB_BACSU	O34878	bacillus su
740	5	3.1	217	1	PSB9_ONCMY	Q9pt26	oncorhynchu
741	5	3.1	217	1	PSB9_ORYLA	Q8uw64	oryzias lat
742	5	3.1	217	1	R10A_HUMAN	P53025	homo sapien
743	5	3.1	217	1	R10A_MOUSE	P53026	mus musculu
744	5	3.1	217	1	RHRE_PEA	Q9s8p4	pisum sativ
745	5	3.1	217	1	YL83_ARCFU	O28100	archaeoglob
746	5	3.1	217	1	YNJD_ECOLI	P76909	escherichia
747	5	3.1	218	1	CD28_RAT	P31042	rattus norv
748	5	3.1	218	1	KAD_BORPE	P39068	bordetella
749	5	3.1	218	1	R27B_HUMAN	O00194	homo sapien
750	5	3.1	218	1	Y232_SYNY3	Q55705	synechocyst
751	5	3.1	218	1	Y522_HAEIN	Q57256	haemophilus
752	5	3.1	218	1	YIDX_ECOLI	P31461	escherichia
753	5	3.1	218	1	YT44_CAEEL	Q11072	caenorhabdi
754	5	3.1	219	1	PAA4_ECOLI	P22996	escherichia
755	5	3.1	219	1	PSB9_MOUSE	P28076	mus musculu
756	5	3.1	219	1	PSB9_MUSDU	O35521	mus dunni (
757	5	3.1	219	1	PSB9_MUSMB	O35522	mus musculu
758	5	3.1	219	1	PSB9_MUSPL	O35523	mus platyth
759	5	3.1	219	1	PSB9_MUSSI	O35524	mus spicile
760	5	3.1	219	1	RHIB_RHILV	Q03314	rhizobium l
761	5	3.1	219	1	UPP_SCHPO	O13867	schizosacch
762	5	3.1	219	1	VG11_BPT4	P10929	bacteriopha
763	5	3.1	220	1	HIS7_YEAST	P06633	saccharomyc
764	5	3.1	220	1	Y085_MYCTU	Q10882	mycobacteri
765	5	3.1	220	1	Y21A_MYCPN	P75482	mycoplasma
766	5	3.1	221	1	VA36_VACCV	P21059	vaccinia vi
767	5	3.1	221	1	Y310_ARCFU	O29934	archaeoglob
768	5	3.1	221	1	YAC9_STRCO	Q04310	streptomyce
769	5	3.1	221	1	YZ18_AQUAE	O66409	aquifex aeo
770	5	3.1	222	1	KAD_RALSO	Q8xwe1	ralstonia s
771	5	3.1	222	1	TYSY_METJA	Q57931	methanococc
772	5	3.1	223	1	CTL4_MOUSE	P09793	mus musculu
773	5	3.1	223	1	HIS7_CANAL	P56090	candida alb
774	5	3.1	223	1	MOXX_PARDE	P29904	paracoccus
775	5	3.1	223	1	Y117_MYCPN	P75518	mycoplasma
776	5	3.1	223	1	YB43_YEAST	P38304	saccharomyc
777	5	3.1	224	1	GTXA_ARATH	P46421	arabidopsis
778	5	3.1	224	1	HB2S_HUMAN	P01916	homo sapien
779	5	3.1	224	1	HEX8_ADECU	Q90097	canine aden
780	5	3.1	224	1	MTGA_ACICA	O24849	acinetobact
781	5	3.1	224	1	YMFK_ECOLI	P75974	escherichia
782	5	3.1	225	1	GIDB_THETH	Q91cy2	thermus the
783	5	3.1	225	1	KDPE_ECOLI	P21866	escherichia
784	5	3.1	225	1	MTRD_METKA	O32864	methanopyru
785	5	3.1	226	1	HEM4_ARCFU	O30120	archaeoglob
786	5	3.1	227	1	RHON_HUMAN	P52198	homo sapien
787	5	3.1	227	1	RHON_MOUSE	Q9qym5	mus musculu
788	5	3.1	228	1	AQPZ_BRUAB	Q91a79	brucella ab
789	5	3.1	228	1	AQPZ_BRUME	Q91772	brucella me
790	5	3.1	228	1	YAZ3_SCHPO	Q10224	schizosacch
791	5	3.1	228	1	YHY1_YEAST	P38869	saccharomyc
792	5	3.1	228	1	YWS1_CAEEL	Q10937	caenorhabdi
793	5	3.1	229	1	CMF3_BACSU	P39147	bacillus su
794	5	3.1	229	1	HA23_MOUSE	P14439	mus musculu
795	5	3.1	229	1	HIS4_PYRAE	Q8zy14	pyrobaculum
796	5	3.1	229	1	RS3_ARCFU	O28360	archaeoglob
797	5	3.1	229	1	TRYP_SQUAC	P00764	squalus aca
798	5	3.1	229	1	VP25_NPVOP	Q06906	orgyia pseu
799	5	3.1	229	1	VP8_BTV2A	Q04687	bluetongue
800	5	3.1	230	1	COX2_POLOR	Q96183	polypterus
801	5	3.1	230	1	RK1_ODOSI	P49544	odontella s
802	5	3.1	231	1	KAD_PRUAR	O24464	prunus arme
803	5	3.1	231	1	MINC_ECOLI	P18196	escherichia
804	5	3.1	231	1	V054_FOWPV	P21970	fowlpox vir
805	5	3.1	231	1	Y236_CAMJE	Q9piq8	campylobact
806	5	3.1	231	1	YEJ4_YEAST	P40008	saccharomyc
807	5	3.1	232	1	6PGL_CAUCR	Q9a6n1	caulobacter
808	5	3.1	232	1	6PGL_RHIME	Q9z3s1	rhizobium m
809	5	3.1	232	1	FLPA_SULSO	P58032	sulfolobus
810	5	3.1	232	1	FRE_SALTY	Q91619	salmonella
811	5	3.1	232	1	HIS7_SACKL	Q02986	saccharomyc
812	5	3.1	232	1	HYEP_STIAU	Q06816	stigmatella
813	5	3.1	232	1	MTN_ECOLI	P24247	escherichia
814	5	3.1	232	1	Y685_METTH	O26781	methanobact
815	5	3.1	232	1	Y836_AQUAE	O67008	aquifex aeo
816	5	3.1	232	1	YNF7_RHIME	Q52992	rhizobium m
817	5	3.1	233	1	LYSE_CORGL	P94633	corynebacte
818	5	3.1	233	1	SNG4_MOUSE	Q9z112	mus musculu
819	5	3.1	233	1	SODM_YEAST	P00447	saccharomyc
820	5	3.1	233	1	TL16_ARATH	O22773	arabidopsis
821	5	3.1	233	1	TPL_TREPA	P45685	treponema p
822	5	3.1	233	1	YCBG_BACSU	P42239	bacillus su
823	5	3.1	234	1	PSB6_TOBAC	P93395	nicotiana t
824	5	3.1	234	1	Y523_METJA	Q57943	methanococc
825	5	3.1	234	1	YA78_METJA	Q58478	methanococc
826	5	3.1	234	1	YNK0_YEAST	P50945	saccharomyc
827	5	3.1	235	1	COLI_RAT	P01194	r corticotr
828	5	3.1	235	1	COMB_CLOAB	Q97e82	clostridium
829	5	3.1	235	1	FLPA_PYRAE	Q8zti9	pyrobaculum
830	5	3.1	235	1	MOR4_HUMAN	Q9y690	homo sapien
831	5	3.1	235	1	PYRF_AGRT5	P58638	agrobacteri
832	5	3.1	235	1	RS4E_THEAC	Q56230	thermoplasm
833	5	3.1	235	1	RS4E_THEVO	Q97bw4	thermoplasm
834	5	3.1	235	1	TRYD_HUMAN	Q9bzj3	homo sapien
835	5	3.1	235	1	YC15_HAEIN	P44128	haemophilus
836	5	3.1	236	1	ATP6_BACME	P20600	bacillus me

837	5	3.1	236	1	YH76_AQUAE	O67652 aquifex aeo
838	5	3.1	237	1	DUP1_YEAST	P53174 saccharomyc
839	5	3.1	237	1	LECA_DIOGR	P08902 dioclea gra
840	5	3.1	237	1	PSB6_RAT	P28073 rattus norv
841	5	3.1	237	1	PYRE_RALSO	Q8y342 ralstonia s
842	5	3.1	237	1	RK1_CYAPA	P48125 cyanophora
843	5	3.1	237	1	RL2_ARCFU	O28357 archaeoglob
844	5	3.1	237	1	UREF_STRSL	Q55056 streptococc
845	5	3.1	237	1	YAAW_ECO57	P58316 escherichia
846	5	3.1	238	1	PSB6_MOUSE	Q60692 mus musculu
847	5	3.1	238	1	Y035_TREPA	O83078 treponema p
848	5	3.1	239	1	DHAP_BOVIN	P30907 bos taurus
849	5	3.1	239	1	PDXJ_ANASP	Q8z080 anabaena sp
850	5	3.1	239	1	PSB6_HUMAN	P28072 homo sapien
851	5	3.1	239	1	PYRF_BACSU	P25971 bacillus su
852	5	3.1	239	1	RL2_HALMA	P20276 haloarcula
853	5	3.1	239	1	SFSA_AGRT5	P58429 agrobacteri
854	5	3.1	239	1	UL20_HSVEB	P28971 equine herp
855	5	3.1	239	1	YEHC_ECOLI	P33342 escherichia
856	5	3.1	239	1	YN82_VIBCH	Q9kpi5 vibrio chol
857	5	3.1	240	1	ISPF_PLAFD	O96178 plasmodium
858	5	3.1	240	1	YA02_SCHPO	Q09675 schizosacch
859	5	3.1	240	1	YEEI_BACSU	O31509 bacillus su
860	5	3.1	241	1	CRTA_RHOCA	P17055 rhodobacter
861	5	3.1	241	1	HIS4_THEMA	Q9x0c7 thermotoga
862	5	3.1	241	1	MLTE_ECOLI	P76009 escherichia
863	5	3.1	241	1	RT10_ARATH	P42797 arabidopsis
864	5	3.1	241	1	SPIR_SPIME	P21625 spiroplasma
865	5	3.1	241	1	Y513_METJA	Q57933 methanococc
866	5	3.1	242	1	PDXJ_ECOLI	P24223 escherichia
867	5	3.1	242	1	PDXJ_SALTI	Q8z4k6 salmonella
868	5	3.1	242	1	PDXJ_SALTY	Q8zn19 salmonella
869	5	3.1	242	1	PDXJ_SYNY3	P72776 synechocyst
870	5	3.1	242	1	RSTA_ECOLI	P52108 escherichia
871	5	3.1	243	1	PDXJ_YERPE	Q8zcp4 yersinia pe
872	5	3.1	243	1	YO27_CAEEL	P34677 caenorhabdi
873	5	3.1	244	1	CLDC_HUMAN	P56749 homo sapien
874	5	3.1	244	1	LEC_LATSP	P16349 lathyrus sp
875	5	3.1	244	1	UL04_VZVD	P09304 varicella-z
876	5	3.1	244	1	VSID_BPP4	P05461 bacteriopha
877	5	3.1	244	1	YK91_MYCTU	Q10700 mycobacteri
878	5	3.1	245	1	ERM1_ENTFA	P12038 enterococcu
879	5	3.1	245	1	ERM2_ENTFA	P20173 enterococcu
880	5	3.1	245	1	ERMB_ECOLI	P10738 escherichia
881	5	3.1	245	1	ERM_STRPN	P21236 streptococc
882	5	3.1	245	1	ERM_STRSA	P06573 streptococc
883	5	3.1	245	1	GILX_HELHO	P43685 heloderma h
884	5	3.1	245	1	TATC_CAMJE	Q9pht8 campylobact
885	5	3.1	245	1	Y474_TREPA	O83487 treponema p
886	5	3.1	246	1	FABG_THEMA	Q9x248 thermotoga
887	5	3.1	246	1	MCT2_SHEEP	P79204 ovis aries
888	5	3.1	246	1	MOG_BOVIN	P55803 bos taurus
889	5	3.1	247	1	AFAB_ECOLI	P53516 escherichia
890	5	3.1	247	1	NFAE_ECOLI	P46738 escherichia
891	5	3.1	247	1	PS72_XENLA	Q9pvq1 xenopus lae
892	5	3.1	247	1	RECO_BRUME	Q8yg76 brucella me
893	5	3.1	247	1	TRY2_HUMAN	P07478 homo sapien
894	5	3.1	247	1	YJ83_YEAST	P47150 saccharomyc
895	5	3.1	248	1	AMPM_BACHD	Q9z9j4 bacillus ha
896	5	3.1	248	1	IFE3_CAEEL	O61955 caenorhabdi
897	5	3.1	248	1	PS71_XENLA	Q9pvy6 xenopus lae
898	5	3.1	248	1	PSA7_HUMAN	O14818 homo sapien
899	5	3.1	248	1	PSA7_MOUSE	Q9z2u0 mus musculu
900	5	3.1	248	1	SDGF_MOUSE	P31955 mus musculu
901	5	3.1	248	1	SRP_VITRI	Q9sw70 vitis ripar
902	5	3.1	248	1	TRPA_ARCFU	O28673 archaeoglob
903	5	3.1	248	1	Y432_LISMO	P25145 listeria mo
904	5	3.1	248	1	Y452_LISIN	Q92ek7 listeria in
905	5	3.1	249	1	HIS6_SULTO	Q970z0 sulfolobus
906	5	3.1	249	1	PS71_DROME	P22769 drosophila
907	5	3.1	249	1	PSA7_CHICK	O13268 gallus gall
908	5	3.1	249	1	RM11_YEAST	P36521 saccharomyc
909	5	3.1	249	1	Y084_METJA	Q60392 methanococc
910	5	3.1	249	1	Y475_CHLMU	Q9pkj0 chlamydia m
911	5	3.1	249	1	YS91_MYCTU	Q10812 mycobacteri
912	5	3.1	250	1	FCG3_BOVIN	P79107 bos taurus
913	5	3.1	250	1	HCYC_SEPOF	P56824 sepia offic
914	5	3.1	250	1	Y938_METJA	Q58348 methanococc
915	5	3.1	250	1	YJ72_CORGL	P40111 corynebacte
916	5	3.1	251	1	HIS6_LISIN	Q92e88 listeria in
917	5	3.1	251	1	HIS6_LISMO	Q8y9g5 listeria mo
918	5	3.1	251	1	HIS6_STRCO	Q9s2t7 streptomyce
919	5	3.1	251	1	HIS6_SULSO	O33774 sulfolobus
920	5	3.1	251	1	KLKE_HUMAN	Q9p0g3 homo sapien
921	5	3.1	251	1	TRPA_PYRKO	Q9yga9 pyrococcus
922	5	3.1	251	1	Y116_MYCGE	P47362 mycoplasma
923	5	3.1	251	1	YAT8_SCHPO	Q10153 schizosacch
924	5	3.1	252	1	HIS6_STAAM	Q99qw8 staphylococ
925	5	3.1	252	1	OLP1_LYCES	Q41350 lycopersico
926	5	3.1	252	1	TFXE_RHILT	P42727 rhizobium l
927	5	3.1	252	1	Y455_RICPR	Q9zd88 rickettsia
928	5	3.1	253	1	ATP6_CHOCR	P48878 chondrus cr
929	5	3.1	253	1	CAC3_BOVIN	P05805 bos taurus
930	5	3.1	253	1	ENVY_ECOLI	P10805 escherichia
931	5	3.1	253	1	HIS6_CLOAB	Q97kh8 clostridium
932	5	3.1	253	1	HIS6_THETN	Q8r885 thermoanaer
933	5	3.1	253	1	NIKD_ECOLI	P33593 escherichia
934	5	3.1	253	1	RECO_RHILO	Q985a3 rhizobium l
935	5	3.1	253	1	VIRG_AGRT5	P07545 agrobacteri
936	5	3.1	253	1	VP8_FOWPV	P15913 fowlpox vir
937	5	3.1	253	1	YBFH_APLCA	P41824 aplysia cal
938	5	3.1	254	1	HA2R_HUMAN	P01903 homo sapien
939	5	3.1	254	1	IF2A_THEAC	Q9hix3 thermoplasm
940	5	3.1	254	1	PSA7_RAT	P48004 rattus norv
941	5	3.1	255	1	GLIC_MOUSE	Q9esn4 mus musculu
942	5	3.1	255	1	HA21_MOUSE	P01904 mus musculu
943	5	3.1	255	1	HA22_MOUSE	P04224 mus musculu
944	5	3.1	255	1	HCD2_DROME	O18404 drosophila
945	5	3.1	255	1	HIS6_NEIMA	Q9jvh5 neisseria m
946	5	3.1	255	1	HIS6_NEIMB	Q9k0h4 neisseria m
947	5	3.1	255	1	PLSC_NEIGO	Q59601 neisseria g
948	5	3.1	255	1	PLSC_NEIMA	Q9ju41 neisseria m
949	5	3.1	255	1	PLSC_NEIMB	Q9jz47 neisseria m
950	5	3.1	255	1	TNR9_HUMAN	Q07011 homo sapien
951	5	3.1	255	1	TRY4_LUCCU	P35044 lucilia cup
952	5	3.1	256	1	BIOH_ECOLI	P13001 escherichia
953	5	3.1	256	1	HI61_PSEAE	Q9hu44 pseudomonas
954	5	3.1	256	1	HIS6_RALSO	Q8xv85 ralstonia s
955	5	3.1	256	1	PSB5_CHICK	P34065 gallus gall
956	5	3.1	256	1	PTMA_CAMCO	Q45983 campylobact
957	5	3.1	256	1	TRYB_MANSE	P35046 manduca sex
958	5	3.1	256	1	TRYE_DROME	P35005 drosophila
959	5	3.1	256	1	VBR1_CLVN	P14970 cassava lat
960	5	3.1	257	1	ARGB_YERPE	Q8za87 yersinia pe
961	5	3.1	257	1	CBIF_SALTY	Q05630 salmonella
962	5	3.1	257	1	VSP2_TRIMU	Q91508 trimeresuru
963	5	3.1	257	1	VSP4_TRIMU	Q91510 trimeresuru
964	5	3.1	257	1	VSP5_TRIMU	Q91511 trimeresuru
965	5	3.1	257	1	VSP7_TRIMU	Q9dg84 trimeresuru
966	5	3.1	258	1	HB2P_HUMAN	P04440 homo sapien
967	5	3.1	258	1	HB2Q_HUMAN	P13763 homo sapien
968	5	3.1	258	1	HIS6_CORGL	O31139 corynebacte
969	5	3.1	258	1	PSBP_WHEAT	Q00434 triticum ae
970	5	3.1	258	1	YL48_MYCTU	O06228 mycobacteri
971	5	3.1	259	1	GP5D_CHLPS	Q46263 chlamydia p
972	5	3.1	259	1	HIS6_ANASP	Q8yt31 anabaena sp
973	5	3.1	259	1	YBBM_ECOLI	P77307 escherichia
974	5	3.1	260	1	AG84_MYCTU	P46816 mycobacteri
975	5	3.1	260	1	COLJ_XENLA	P06299 xenopus lae
976	5	3.1	260	1	NQRC_PASMU	Q9cla9 pasteurella
977	5	3.1	260	1	STCE_EMENI	Q00674 emericella
978	5	3.1	261	1	CYSE_BUCAP	P32003 buchnera ap
979	5	3.1	261	1	DHG_BACSU	P12310 bacillus su
980	5	3.1	261	1	HIS4_YEAST	P40545 saccharomyc
981	5	3.1	261	1	HIS6_SYNY3	P74106 synechocyst
982	5	3.1	261	1	SMTA_ECOLI	P36566 escherichia

983	5	3.1	261	1	SURE_METKA	Q8ty72 methanopyru
984	5	3.1	261	1	YH86_SYNY3	P73335 synechocyst
985	5	3.1	261	1	ZNUB_ECOLI	P39832 escherichia
986	5	3.1	262	1	DHSB_CYACA	P48933 cyanidium c
987	5	3.1	262	1	FBOL_BOMMO	P21828 bombyx mori
988	5	3.1	262	1	ITMB_CHICK	O42204 gallus gall
989	5	3.1	262	1	LAT_HUMAN	O43561 homo sapien
990	5	3.1	262	1	TRUA_PYRHO	O58941 pyrococcus
991	5	3.1	262	1	VG38_BPT2	P07875 bacteriopha
992	5	3.1	262	1	Y096_METJA	Q57561 methanococc
993	5	3.1	262	1	YA23_METJA	Q58429 methanococc
994	5	3.1	262	1	YLU4_CAEEL	P34398 caenorhabdi
995	5	3.1	263	1	MPP1_PHLPR	P43213 phleum prat
996	5	3.1	263	1	SPAR_SALTY	P40701 salmonella
997	5	3.1	264	1	CTD2_RAT	O35116 rattus norv
998	5	3.1	264	1	XYLI_PSEPU	P49155 pseudomonas
999	5	3.1	264	1	Y902_HAEIN	P44070 haemophilus
1000	5	3.1	265	1	HME2_BRARE	P09015 brachydanio

ALIGNMENTS

```
RESULT 1
FIMH_ECOLI
ID   FIMH_ECOLI     STANDARD;      PRT;   300 AA.
AC   P08191;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   FimH protein precursor.
GN   FIMH OR B4320.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88038337; PubMed=2890081;
RA   Klemm P., Christiansen G.;
RT   "Three fim genes required for the regulation of length and mediation
RT   of adhesion of Escherichia coli type 1 fimbriae.";
RL   Mol. Gen. Genet. 208:439-445(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=95334362; PubMed=7610040;
RA   Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA   Blattner F.R.;
RT   "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT   region from 92.8 through 100 minutes.";
RL   Nucleic Acids Res. 23:2105-2119(1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=90256291; PubMed=1971261;
RA   Krogfelt K.A., Bergmans H., Klemm P.;
RT   "Direct evidence that the FimH protein is the mannose-specific
RT   adhesin of Escherichia coli type 1 fimbriae.";
RL   Infect. Immun. 58:1995-1998(1990).
CC   -!- FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
CC       ADHESION OF TYPE 1 FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION
CC       OF FIMBRIAE). ADHESIN RESPONSIBLE FOR THE BINDING TO D-MANNOSE. IT
CC       IS LATERALLY POSITIONED AT INTERVALS IN THE STRUCTURE OF THE TYPE
CC       1 FIMBRIAE. IN ORDER TO INTEGRATE FIMH IN THE FIMBRIAE FIMF AND
CC       FIMG ARE NEEDED.
CC   -!- SIMILARITY: THIS PROTEIN EXHIBITS HOMOLOGY WITH THE FIMBRIAL
CC       SUBUNIT PROTEIN FIMA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X05672; CAA29156.1; -.
DR   EMBL; U14003; AAA97216.1; -.
DR   EMBL; AE000502; AAC77276.1; -.
DR   PIR; S09563; S09563.
DR   EcoGene; EG10315; fimH.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
KW   Fimbria; Signal; Complete proteome.
FT   SIGNAL       1     23       POTENTIAL.
FT   CHAIN       24    300       FIMH PROTEIN.
FT   CONFLICT   197    197       P -> R (IN REF. 1).
FT   CONFLICT   222    222       T -> H (IN REF. 1).
SQ   SEQUENCE   300 AA;  31473 MW;  939204A51658747D CRC64;

  Query Match             57.1%;  Score 92;  DB 1;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-89;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 2
YDEQ_ECOLI
ID   YDEQ_ECOLI     STANDARD;      PRT;   304 AA.
AC   P77588;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical fimbrial-like protein ydeQ precursor.
GN   YDEQ OR B1502.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
CC   -!- SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE000247; AAC74575.1; -.
DR   EMBL; D90792; BAA15175.1; -.
DR   EMBL; D90793; BAA15183.1; -.
DR   EcoGene; EG13799; ydeQ.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
KW   Hypothetical protein; Fimbria; Signal; Complete proteome.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    304       HYPOTHETICAL FIMBRIAL-LIKE PROTEIN YDEQ.
SQ   SEQUENCE   304 AA;  32069 MW;  8153C86E3087D99A CRC64;

  Query Match             5.6%;  Score 9;  DB 1;  Length 304;
  Best Local Similarity   100.0%;  Pred. No. 0.096;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 GQNLVVDLS 14
           |||||||||
Db      53 GQNLVVDLS 61


RESULT 3
RL34_ARATH
ID   RL34_ARATH     STANDARD;      PRT;   120 AA.
AC   Q42351;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   60S ribosomal protein L34.
GN   RPL34 OR AT1G26880 OR T2P11.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Cooke R., Laudie M., Raynal M., Delseny M.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
CC   -!- SIMILARITY: BELONGS TO THE L34E FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; F20073; CAA23390.1; -.
DR   EMBL; AC005508; AAD14494.1; -.
DR   InterPro; IPR001284; Ribosomal_L34E.
DR   Pfam; PF01199; Ribosomal_L34e; 1.
DR   PRINTS; PR01250; RIBOSOMALL34.
DR   ProDom; PD005148; Ribosomal_L34E; 1.
DR   PROSITE; PS01145; RIBOSOMAL_L34E; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   120 AA;  13705 MW;  5B2334C5C22A44C3 CRC64;

  Query Match             4.3%;  Score 7;  DB 1;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 AYGGVLS 44
           |||||||
Db      77 AYGGVLS 83


RESULT 4
RL34_PEA
ID   RL34_PEA       STANDARD;      PRT;   120 AA.
AC   P40590;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   60S ribosomal protein L34.
GN   RPL34.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Alaska;
RX   MEDLINE=96046745; PubMed=7579177;
RA   Devitt M.L., Stafstrom J.P.;
RT   "Cell cycle regulation during growth-dormancy cycles in pea axillary
RT   buds.";
RL   Plant Mol. Biol. 29:255-265(1995).
CC   -!- SIMILARITY: BELONGS TO THE L34E FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U10047; AAA86953.1; -.
DR   InterPro; IPR001284; Ribosomal_L34E.
DR   Pfam; PF01199; Ribosomal_L34e; 1.
DR   PRINTS; PR01250; RIBOSOMALL34.
DR   ProDom; PD005148; Ribosomal_L34E; 1.
DR   PROSITE; PS01145; RIBOSOMAL_L34E; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   120 AA;  13822 MW;  822C8A93589E15ED CRC64;

  Query Match             4.3%;  Score 7;  DB 1;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 AYGGVLS 44
           |||||||
Db      77 AYGGVLS 83


RESULT 5
RL34_TOBAC
```

```
ID   RL34_TOBAC     STANDARD;      PRT;   120 AA.
AC   P41098;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   60S ribosomal protein L34.
GN   RPL34.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94355650; PubMed=8075394;
RA   Gao J., Kim S.R., Chung Y.Y., Lee J.M., An G.;
RT   "Developmental and environmental regulation of two ribosomal protein
RT   genes in tobacco.";
RL   Plant Mol. Biol. 25:761-770(1994).
CC   -!- SIMILARITY: BELONGS TO THE L34E FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L27107; AAA57159.1; -.
DR   EMBL; L27089; AAA57158.1; -.
DR   InterPro; IPR001284; Ribosomal_L34E.
DR   Pfam; PF01199; Ribosomal_L34e; 1.
DR   PRINTS; PR01250; RIBOSOMALL34.
DR   ProDom; PD005148; Ribosomal_L34E; 1.
DR   PROSITE; PS01145; RIBOSOMAL_L34E; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   120 AA;  13752 MW;  5A987F20093EF6E9 CRC64;

  Query Match             4.3%;  Score 7;  DB 1;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 5.4;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 AYGGVLS 44
           |||||||
Db      77 AYGGVLS 83


RESULT 6
MYP2_HUMAN
ID   MYP2_HUMAN     STANDARD;      PRT;   131 AA.
AC   P02689;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myelin P2 protein.
GN   PMP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92068191; PubMed=1720307;
RA   Hayasaka K., Nanao K., Tahara M., Sato W., Takada G., Miura M.,
RA   Uyemura K.;
RT   "Isolation and sequence determination of cDNA encoding P2 protein of
RT   human peripheral myelin.";
RL   Biochem. Biophys. Res. Commun. 181:204-207(1991).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE=83058785; PubMed=6183401;
RA   Suzuki M., Kitamura K., Sakamoto Y., Uyemura K.;
RT   "The complete amino acid sequence of human P2 protein.";
RL   J. Neurochem. 39:1759-1762(1982).
RN   [3]
RP   SEQUENCE OF 1-115 FROM N.A.
RX   MEDLINE=95054012; PubMed=7525873;
RA   Narayanan V., Ripepi B., Jabs E.W., Hawkins A., Griffin C.,
RA   Tennekoon G.;
RT   "Partial structure and mapping of the human myelin P2 protein gene.";
RL   J. Neurochem. 63:2010-2013(1994).
CC   -!- FUNCTION: THIS PROTEIN MAY BE A LIPID TRANSPORT PROTEIN IN SCHWANN
CC       CELLS.
CC   -!- MISCELLANEOUS: P2 PROTEIN AND MYELIN BASIC PROTEIN TOGETHER
CC       CONSTITUTE A MAJOR FRACTION OF PERIPHERAL NERVOUS SYSTEM MYELIN
CC       PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE FABP/P2/CRBP/CRABP FAMILY OF
CC       TRANSPORTERS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D16181; BAA03726.1; -.
DR   EMBL; D16179; BAA03726.1; JOINED.
DR   EMBL; D16180; BAA03726.1; JOINED.
DR   EMBL; X62167; CAA44096.1; -.
DR   EMBL; S73470; AAB32592.2; -.
DR   EMBL; S73468; AAB32592.2; JOINED.
DR   EMBL; S73469; AAB32592.2; JOINED.
DR   PIR; A03143; MPHU2.
DR   PIR; JT0977; JT0977.
DR   HSSP; P02690; 1PMP.
DR   Genew; HGNC:9117; PMP2.
DR   MIM; 170715; -.
DR   InterPro; IPR000463; Fatty_acid_BP.
DR   InterPro; IPR000566; Lipocln_cytFABP.
DR   Pfam; PF00061; lipocalin; 1.
DR   PRINTS; PR00178; FATTYACIDBP.
DR   PROSITE; PS00214; FABP; 1.
KW   Myelin; Lipid-binding; Transport; Acetylation.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION.
FT   DISULFID    117    124
FT   CONFLICT     24     24       G -> GG (IN REF. 3).
FT   CONFLICT     98     98       D -> N (IN REF. 2).
FT   CONFLICT    110    110       N -> D (IN REF. 2).
SQ   SEQUENCE   131 AA;  14778 MW;  3D0E640BA03E69D2 CRC64;

  Query Match             4.3%;  Score 7;  DB 1;  Length 131;
  Best Local Similarity   100.0%;  Pred. No. 5.8;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 VTLQRGS 37
           |||||||
Db      84 VTLQRGS 90


RESULT 7
YQHO_BACSU
ID   YQHO_BACSU     STANDARD;      PRT;   291 AA.
AC   P54513;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein yqhO.
GN   YQHO.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE=97124195; PubMed=8969508;
RA   Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA   Kobayashi Y.;
RT   "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT   the Bacillus subtilis genome containing the skin element and many
RT   sporulation genes.";
RL   Microbiology 142:3103-3111(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D84432; BAA12552.1; -.
DR   EMBL; Z99116; CAB14382.1; -.
DR   SubtiList; BG11703; yqhO.
DR   InterPro; IPR002641; Patatin.
DR   Pfam; PF01734; Patatin; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     34     50       POTENTIAL.
SQ   SEQUENCE   291 AA;  32860 MW;  B0A7C5DDE4AE37B8 CRC64;

  Query Match             4.3%;  Score 7;  DB 1;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     40 GGVLSNF 46
          |||||||
Db    184 GGVLSNF 190


RESULT 8
YAS4_SCHPO
ID   YAS4_SCHPO     STANDARD;      PRT;   338 AA.
AC   Q10140;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein C3H8.04 in chromosome I.
GN   SPAC3H8.04.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69086; CAA93161.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   338 AA;  38474 MW;  5BB40B5117F9EF74 CRC64;

  Query Match             4.3%;  Score 7;  DB 1;  Length 338;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    104 LILRQTN 110
          |||||||
Db    237 LILRQTN 243


RESULT 9
CTR1_HUMAN
ID   CTR1_HUMAN     STANDARD;      PRT;   629 AA.
AC   P30825;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   High-affinity cationic amino acid transporter-1 (CAT-1) (CAT1) (System
DE   Y+ basic amino acid transporter) (Ecotropic retroviral leukemia
```

```
DE   receptor homolog) (ERR) (Ecotropic retrovirus receptor homolog).
GN   SLC7A1 OR ATRC1 OR REC1L.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92024065; PubMed=1718082;
RA   Yoshimoto T., Yoshimoto E., Meruelo D.;
RT   "Molecular cloning and characterization of a novel human gene
RT   homologous to the murine ecotropic retroviral receptor.";
RL   Virology 185:10-17(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Urinary bladder;
RX   MEDLINE=92217962; PubMed=1348489;
RA   Albritton L.M., Bowcock A.M., Eddy R.L., Morton C.C., Tseng L.,
RA   Farrer L.A., Cavalli-Sforza L.L., Shows T.B., Cunningham J.M.;
RT   "The human cationic amino acid transporter (ATRC1): physical and
RT   genetic mapping to 13q12-q14.";
RL   Genomics 12:430-434(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Kamath S.G., Furesz T.C., Way B.A., Smith C.H.;
RT   "Cloning, expression, and characterization of a cationic amino acid
RT   transporter from cultured human placental trophoblast cells.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HIGH-AFFINITY, LOW CAPACITY PERMEASE INVOLVED IN THE
CC       TRANSPORT OF THE CATIONIC AMINO ACIDS (ARGININE, LYSINE AND
CC       ORNITHINE) IN NON-HEPATIC TISSUES. MAY ALSO FUNCTION AS AN
CC       ECOTROPIC RETROVIRAL LEUKEMIA RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC   -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MANY AMINO ACID
CC       PERMEASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59155; CAA41869.1; -.
DR   EMBL; X57303; CAA40560.1; -.
DR   EMBL; AF078107; AAC27721.1; -.
DR   Genew; HGNC:11057; SLC7A1.
DR   MIM; 104615; -.
DR   InterPro; IPR002293; AA/rel_prmease1.
DR   InterPro; IPR004755; Cat_AA_permease.
DR   InterPro; IPR004841; Permease.
DR   Pfam; PF00324; aa_permeases; 1.
DR   TIGRFAMs; TIGR00906; 2A0303; 1.
KW   Transmembrane; Glycoprotein; Transport; Amino-acid transport;
KW   Receptor.
FT   DOMAIN       1     35       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     36    57       POTENTIAL.
FT   DOMAIN       58    61       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     62    82       POTENTIAL.
FT   DOMAIN       83    102      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     103   123      POTENTIAL.
FT   DOMAIN       124   162      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     163   183      POTENTIAL.
FT   DOMAIN       184   191      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     192   212      POTENTIAL.
FT   DOMAIN       213   246      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     247   267      POTENTIAL.
FT   DOMAIN       268   287      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     288   307      POTENTIAL.
FT   DOMAIN       308   337      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     338   358      POTENTIAL.
FT   DOMAIN       359   384      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     385   405      POTENTIAL.
FT   DOMAIN       406   408      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     409   429      POTENTIAL.
FT   DOMAIN       430   492      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     493   513      POTENTIAL.
FT   DOMAIN       514   526      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     527   551      POTENTIAL.
FT   DOMAIN       552   559      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     560   580      POTENTIAL.
FT   DOMAIN       581   584      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     585   605      POTENTIAL.
FT   DOMAIN       606   629      CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     226   226      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     235   235      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      23    23      R -> P (IN REF. 2).
SQ   SEQUENCE   629 AA;  67638 MW;  717734D4793647C5 CRC64;

  Query Match              4.3%;  Score 7;  DB 1;  Length 629;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 SLIAVLI 105
           |||||||
Db     501 SLIAVLI 507

RESULT 10
YBHJ_ECOLI
ID   YBHJ_ECOLI     STANDARD;      PRT;   753 AA.
AC   P75764; Q9R7S5;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ybhJ.
GN   YBHJ OR B0771.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
CC   -!- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; AE000179; AAC73858.1; ALT_INIT.
DR   EMBL; D90715; BAA35435.1; ALT_INIT.
DR   EcoGene; EG13666; ybhJ.
DR   InterPro; IPR001030; Aconitase_N.
DR   Pfam; PF00330; aconitase; 1.
DR   ProDom; PD000511; Aconitase_N; 1.
DR   PROSITE; PS00450; ACONITASE_1; FALSE_NEG.
DR   PROSITE; PS01244; ACONITASE_2; FALSE_NEG.
KW   Hypothetical protein; Iron-sulfur; Complete proteome.
FT   METAL       360    360       IRON (IRON-SULFUR CLUSTER) (BY
FT                                SIMILARITY).
FT   METAL       421    421       IRON (IRON-SULFUR CLUSTER) (BY
FT                                SIMILARITY).
FT   METAL       424    424       IRON (IRON-SULFUR CLUSTER) (BY
FT                                SIMILARITY).
SQ   SEQUENCE   753 AA;  81504 MW;  A75D1AE0F7CDF588 CRC64;

  Query Match             4.3%;  Score 7;  DB 1;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLI 101
           |||||||
Db     739 IKAGSLI 745


RESULT 11
YEGN_ECOLI
ID   YEGN_ECOLI     STANDARD;      PRT;  1040 AA.
AC   P76398; O08005;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein yegN.
GN   YEGN OR B2075.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251358; PubMed=9097040;
RA   Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA   Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA   Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA   Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA   Sivasundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA   Yamamoto Y., Horiuchi T.;
RT   "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 40.1-50.0 min region on the linkage map.";
RL   DNA Res. 3:379-392(1996).
CC   -!- FUNCTION: COULD BE A DRUG EFFLUX PUMP (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE ACRB/ACRD/ACRF FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000297; AAC75136.1; -.
DR   EMBL; D90845; BAA15929.1; -.
DR   EMBL; D90846; BAA15931.1; -.
DR   EcoGene; EG14057; yegN.
DR   InterPro; IPR001036; Acrflvin_res.
DR   Pfam; PF00873; ACR_tran; 1.
DR   PRINTS; PR00702; ACRIFLAVINRP.
KW   Hypothetical protein; Transmembrane; Inner membrane; Transport;
KW   Complete proteome.
FT   DOMAIN        1     20       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM     21     39       1 (BY SIMILARITY).
FT   DOMAIN       40    342       PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM    343    362       2 (BY SIMILARITY).
FT   DOMAIN      363    368       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    369    388       3 (BY SIMILARITY).
FT   DOMAIN      389    394       PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM    395    416       4 (BY SIMILARITY).
FT   DOMAIN      417    443       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    444    462       5 (BY SIMILARITY).
FT   DOMAIN      463    475       PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM    476    498       6 (BY SIMILARITY).
FT   DOMAIN      499    536       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    537    555       7 (BY SIMILARITY).
FT   DOMAIN      556    865       PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM    866    885       8 (BY SIMILARITY).
FT   DOMAIN      886    891       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    892    911       9 (BY SIMILARITY).
FT   DOMAIN      912    917       PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM    918    939       10 (BY SIMILARITY).
FT   DOMAIN      940    967       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    968    986       11 (BY SIMILARITY).
FT   DOMAIN      987    999       PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1000   1022       12 (BY SIMILARITY).
FT   DOMAIN     1023   1040       CYTOPLASMIC (BY SIMILARITY).
SQ   SEQUENCE   1040 AA;  112077 MW;  195E3D989C7AC6F0 CRC64;

  Query Match             4.3%;  Score 7;  DB 1;  Length 1040;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 SLIAVLI 105
           |||||||
Db     450 SLIAVLI 456


RESULT 12
VPO_BPPRD
ID   VPO_BPPRD      STANDARD;      PRT;    68 AA.
AC   P27390;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
DE   Protein O (GPO).
GN   O.
OS   Bacteriophage PRD1.
OC   Viruses; dsDNA viruses, no RNA stage; Tectiviridae; Tectivirus.
OX   NCBI_TaxID=10658;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 61-68.
RX   MEDLINE=91306449; PubMed=1853567;
RA   Bamford J.K.H., Haenninen A.-L., Pakula T.M., Ojala P.M.,
RA   Kalkkinen N., Frilander M., Bamford D.H.;
RT   "Genome organization of membrane-containing bacteriophage PRD1.";
RL   Virology 183:658-676(1991).
CC   -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M69077; AAA32470.1; -.
DR   PIR; G36776; WMBPUB.
SQ   SEQUENCE   68 AA;  6695 MW;  86E351A8AF13746A CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 68;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      94 VIKAGS 99
           ||||||
Db      34 VIKAGS 39


RESULT 13
IVB2_BUNMU
ID   IVB2_BUNMU     STANDARD;      PRT;    85 AA.
AC   P00989; O42299;
DT   21-JUL-1986 (Rel. 01, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-2 bungarotoxin B chain precursor (Bungarotoxin, B2 chain).
OS   Bungarus multicinctus (Many-banded krait).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Elapidae; Bungarinae; Bungarus.
OX   NCBI_TaxID=8616;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Venom gland;
RX   MEDLINE=98359733; PubMed=9693106;
RA   Wu P.F., Wu S.N., Chang C.C., Chang L.S.;
RT   "Cloning and functional expression of B chains of beta-bungarotoxins
RT   from Bungarus multicinctus (Taiwan banded krait).";
RL   Biochem. J. 334:87-92(1998).
RN   [2]
RP   SEQUENCE OF 25-85.
RC   TISSUE=Venom;
RX   MEDLINE=82239269; PubMed=7096304;
RA   Kondo K., Toda H., Narita K., Lee C.-Y.;
RT   "Amino acid sequence of beta 2-bungarotoxin from Bungarus
RT   multicinctus venom. The amino acid substitutions in the B chains.";
RL   J. Biochem. 91:1519-1530(1982).
CC   -!- FUNCTION: BETA-2 BUNGAROTOXIN IS A PRESYNAPTIC NEUROTOXIN OF THE
CC       VENOM. THE B CHAIN IS HOMOLOGOUS TO VENOM BASIC PROTEASE
CC       INHIBITORS BUT HAS NO PROTEASE INHIBITOR ACTIVITY AND IS NON-
CC       TOXIC.
CC   -!- SUBUNIT: DIMER OF DISSIMILAR CHAINS LINKED BY A DISULFIDE BOND.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Y12101; CAA72810.1; -.
DR   PIR; A01220; TIKFB2.
DR   HSSP; P00981; 1DTK.
DR   InterPro; IPR002223; Kunitz_BPTI.
DR   Pfam; PF00014; Kunitz_BPTI; 1.
DR   PRINTS; PR00759; BASICPTASE.
DR   ProDom; PD000222; Kunitz_BPTI; 1.
DR   SMART; SM00131; KU; 1.
DR   PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR   PROSITE; PS50279; BPTI_KUNITZ_2; 1.
KW   Venom; Presynaptic neurotoxin; Signal.
FT   SIGNAL       1     24
FT   CHAIN       25     85       BETA-2 BUNGAROTOXIN B CHAIN.
FT   DOMAIN      31     81       BPTI/KUNITZ INHIBITOR.
FT   DISULFID    31     81       BY SIMILARITY.
FT   DISULFID    40     64       BY SIMILARITY.
FT   DISULFID    56     77       BY SIMILARITY.
FT   DISULFID    79     79       INTERCHAIN (WITH AN A CHAIN) (POTENTIAL).
FT   ACT_SITE    41     42       REACTIVE BOND (BY SIMILARITY).
FT   CONFLICT    44     44       MISSING (IN REF. 2).
FT   CONFLICT    65     70       NGNGNH -> DGDHGN (IN REF. 2).
FT   CONFLICT    82     83       LE -> EL (IN REF. 2).
SQ   SEQUENCE   85 AA;  9568 MW;  FE95A59AF92BF2AA CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 85;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      84 LTPVSS 89
           ||||||
Db      19 LTPVSS 24


RESULT 14
YDGC_ECOLI
ID   YDGC_ECOLI     STANDARD;      PRT;   111 AA.
AC   P52110; P77466;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ydgC.
GN   YDGC OR B1607 OR Z2608 OR ECS2313.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RA   Kuempel P.L.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97251357; PubMed=9097039;
RA   Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA   Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA   Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA   Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA   Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA   Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT   "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 28.0-40.1 min region on the linkage map.";
RL   DNA Res. 3:363-377(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
```

```
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [6]
RP   IDENTIFICATION.
RA   Rudd K.E.;
RL   Unpublished observations (JAN-1996).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: TO P.AERUGINOSA GLPM.
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
CC       IN POSITIONS 11 AND 38.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U41101; -; NOT_ANNOTATED_CDS.
DR   EMBL; AE000256; AAC74679.1; -.
DR   EMBL; D90803; BAA15345.1; -.
DR   EMBL; D90804; BAA15355.1; -.
DR   EMBL; AE005384; AAG56594.1; -.
DR   EMBL; AP002558; BAB35736.1; -.
DR   EcoGene; EG13191; ydgC.
KW   Hypothetical protein; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   TRANSMEM     27     47       POTENTIAL.
FT   TRANSMEM     59     79       POTENTIAL.
FT   TRANSMEM     88    108       POTENTIAL.
SQ   SEQUENCE   111 AA;  12323 MW;  8690061D033E5873 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     92 GLVIKA 97
          ||||||
Db      2 GLVIKA 7


RESULT 15
RFA3_YEAST
ID   RFA3_YEAST     STANDARD;      PRT;   122 AA.
AC   P26755;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Replication factor-A protein 3 (RF-A) (Replication protein A 13 kDa
DE   subunit).
GN   RFA3 OR YJL173C OR J0506.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 8-22; 29-46 AND 119-122.
RC   STRAIN=W303-1A;
RX   MEDLINE=91357474; PubMed=1885001;
RA   Brill S.J., Stillman B.;
RT   "Replication factor-A from Saccharomyces cerevisiae is encoded by
RT   three essential genes coordinately expressed at S phase.";
RL   Genes Dev. 5:1589-1600(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Loehning C., Mueller C., Freidel K., Ciriacy M.;
RL   Submitted (NOV-1990) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Obermaier B., Piravandi E., Rinke M., Domdey H.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS SINGLE-STRANDED DNA. STIMULATES THE ACTIVITY
CC       OF A COGNATE STRAND EXCHANGE PROTEIN (SEP1). IT COOPERATES WITH
CC       T-AG AND DNA TOPOISOMERASE I TO UNWIND TEMPLATE DNA CONTAINING THE
CC       SIMIAN VIRUS 40 ORIGIN OF DNA REPLICATION.
CC   -!- SUBUNIT: HETEROTRIMER OF 69, 36, AND 13 kDa CHAINS. THE
CC       DNA-BINDING ACTIVITY MAY RESIDE EXCLUSIVELY ON THE 69 kDa SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: The N-terminus is blocked.
CC   -!- SIMILARITY: LIMITED, TO MAMMALIAN RFA3/RPA3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59750; CAA42422.1; -.
DR   EMBL; X56792; CAA40110.1; -.
DR   EMBL; Z49448; CAA89468.1; -.
DR   PIR; C37281; C37281.
DR   PIR; S17022; S17022.
DR   SGD; S0003709; RFA3.
KW   DNA replication; Nuclear protein.
SQ   SEQUENCE   122 AA;  13816 MW;  5EC5D6FB56BAC008 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 122;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     63 SETPRV 68
          ||||||
Db      3 SETPRV 8


RESULT 16
NB5M_CHICK
ID   NB5M_CHICK     STANDARD;      PRT;   133 AA.
AC   P48306;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative NADH-ubiquinone oxidoreductase B15 subunit (EC 1.6.5.3)
DE   (EC 1.6.99.3) (Complex I-B15) (CI-B15) (Hypothetical protein Walter)
DE   (GGHPW).
GN   NDUFB4.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94040816; PubMed=7901127;
RA   Goldberg G.S., Kaczmarczyk W.;
RT   "A chicken genomic DNA fragment that hybridizes to the murine Hox-3.1
RT   homeobox is likely to encode the NADH ubiquinone oxidoreductase
RT   subunit B15.";
RL   Gene 133:233-235(1993).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93077061; PubMed=1359990;
RA   Goldberg G.S., Kaczmarczyk W.;
RT   "Sequence of a novel chicken genomic DNA fragment that hybridizes to
RT   the murine Hox-3.1 homeobox.";
RL   Gene 121:397-398(1992).
CC   -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC       CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC       TO BE UBIQUINONE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC   -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane; matrix side
CC       (By similarity).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60778; CAA43193.1; ALT_SEQ.
DR   EMBL; M84354; AAA70193.1; ALT_SEQ.
KW   Hypothetical protein; Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ   SEQUENCE   133 AA;  14938 MW;  ED7E82948C97B352 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 133;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     87 VSSAGG 92
          ||||||
Db    117 VSSAGG 122


RESULT 17
YEDX_ECO57
ID   YEDX_ECO57     STANDARD;      PRT;   137 AA.
AC   Q8XB75;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Transthyretin-like protein precursor.
GN   YEDX OR Z3062 OR ECS2708.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE TRANSTHYRETIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005418; AAG56983.1; -.
DR   EMBL; AP002559; BAB36131.1; -.
DR   InterPro; IPR000895; Transthyretin.
DR   PROSITE; PS00768; TRANSTHYRETIN_1; 1.
DR   PROSITE; PS00769; TRANSTHYRETIN_2; 1.
KW   Transport; Periplasmic; Signal; Complete proteome.
FT   SIGNAL        1     23       BY SIMILARITY.
FT   CHAIN        24    137       TRANSTHYRETIN-LIKE PROTEIN.
SQ   SEQUENCE   137 AA;  15537 MW;  0A8D534E6A46B9F8 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 137;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    142 DVTVTL 147
          ||||||
Db     44 DVTVTL 49


RESULT 18
YEDX_ECOLI
ID   YEDX_ECOLI     STANDARD;      PRT;   137 AA.
AC   P76341;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Transthyretin-like protein precursor.
GN   YEDX OR B1970.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   PARTIAL SEQUENCE OF N-TERMINUS.
RA   Rudd K.E.;
RL   Unpublished observations (AUG-1999).
CC   -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE TRANSTHYRETIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000288; AAC75036.1; -.
DR   HSSP; P27731; 1TFP.
DR   EcoGene; EG14046; yedX.
DR   InterPro; IPR000895; Transthyretin.
DR   Pfam; PF00576; Transthyretin; 1.
```

```
DR   PRINTS; PR00189; TRNSTHYRETIN.
DR   ProDom; PD003457; Transthyretin; 1.
DR   SMART; SM00095; TR_THY; 1.
DR   PROSITE; PS00768; TRANSTHYRETIN_1; 1.
DR   PROSITE; PS00769; TRANSTHYRETIN_2; 1.
KW   Transport; Periplasmic; Signal; Complete proteome.
FT   SIGNAL        1     23
FT   CHAIN        24    137       TRANSTHYRETIN-LIKE PROTEIN.
SQ   SEQUENCE   137 AA;  15460 MW;  B58A534051DDDC5E CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 137;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    142 DVTVTL 147
          ||||||
Db     44 DVTVTL 49


RESULT 19
RS9_PYRAE
ID   RS9_PYRAE      STANDARD;      PRT;   142 AA.
AC   Q8ZYQ0;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S9P.
GN   RPS9P OR PAE0674.
OS   Pyrobaculum aerophilum.
OC   Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC   Thermoproteaceae; Pyrobaculum.
OX   NCBI_TaxID=13773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX   PubMed=11792869;
RA   Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA   Miller J.H.;
RT   "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT   aerophilum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC   -!- SIMILARITY: BELONGS TO THE S9P FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE009779; AAL62943.1; ALT_INIT.
DR   InterPro; IPR000754; Ribosomal_S9.
DR   Pfam; PF00380; Ribosomal_S9; 1.
DR   ProDom; PD001627; Ribosomal_S9; 1.
DR   PROSITE; PS00360; RIBOSOMAL_S9; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   142 AA;  15930 MW;  8D31075F2E4CFB42 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 142;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     64 ETPRVV 69
          ||||||
Db      9 ETPRVV 14


RESULT 20
H2A2_PEA
ID   H2A2_PEA       STANDARD;      PRT;   149 AA.
AC   P40281;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Histone H2A.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Alaska;
RX   MEDLINE=96046745; PubMed=7579177;
RA   Devitt M.L., Stafstrom J.P.;
RT   "Cell cycle regulation during growth-dormancy cycles in pea axillary
RT   buds.";
RL   Plant Mol. Biol. 29:255-265(1995).
CC   -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC       H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H2A FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U10041; AAA86947.1; -.
DR   InterPro; IPR002119; Histone_H2A.
DR   InterPro; IPR004822; Histone_core.
DR   Pfam; PF00125; histone; 1.
DR   PRINTS; PR00620; HISTONEH2A.
DR   ProDom; PD000522; Histone_H2A; 1.
DR   SMART; SM00414; H2A; 1.
DR   PROSITE; PS00046; HISTONE_H2A; 1.
KW   Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
KW   Multigene family; Acetylation.
FT   MOD_RES       1       1       ACETYLATION (BY SIMILARITY).
SQ   SEQUENCE   149 AA;  15704 MW;  5ED29866408BCE05 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 149;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     38 AYGGVL 43
          ||||||
Db    112 AYGGVL 117


RESULT 21
H2A1_PEA
ID   H2A1_PEA       STANDARD;      PRT;   150 AA.
AC   P25470;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Histone H2A.
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93044503; PubMed=1841722;
RA   Koning A.J., Tanimoto E.Y., Kiehne K., Rost T., Comai L.;
RT   "Cell-specific expression of plant histone H2A genes.";
RL   Plant Cell 3:657-665(1991).
CC   -!- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
CC       H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
```

```
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H2A FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M64838; -; NOT_ANNOTATED_CDS.
DR   PIR; JQ1183; JQ1183.
DR   InterPro; IPR002119; Histone_H2A.
DR   InterPro; IPR004822; Histone_core.
DR   Pfam; PF00125; histone; 1.
DR   PRINTS; PR00620; HISTONEH2A.
DR   ProDom; PD000522; Histone_H2A; 1.
DR   SMART; SM00414; H2A; 1.
DR   PROSITE; PS00046; HISTONE_H2A; 1.
KW   Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
KW   Multigene family; Acetylation.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
SQ   SEQUENCE   150 AA;  15879 MW;  0F72BAFE5265F230 CRC64;

  Query Match              3.7%;  Score 6;  DB 1;  Length 150;
  Best Local Similarity   100.0%;  Pred. No. 74;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 AYGGVL 43
           ||||||
Db     112 AYGGVL 117


RESULT 22
NAPB_RHOSH
ID   NAPB_RHOSH     STANDARD;      PRT;   154 AA.
AC   Q53177;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Diheme cytochrome c napB precursor.
GN   NAPB.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1063;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX   MEDLINE=96332666; PubMed=8730872;
RA   Reyes F., Roldan M.D., Klipp W., Castillo F., Moreno-Vivian C.;
RT   "Isolation of periplasmic nitrate reductase genes from Rhodobacter
RT   sphaeroides DSM 158: structural and functional differences among
RT   prokaryotic nitrate reductases.";
RL   Mol. Microbiol. 19:1307-1318(1996).
CC   -!- FUNCTION: SMALL SUBUNIT OF THE PERIPLASMIC NITRATE REDUCTASE
CC       (NAP). ONLY EXPRESSED AT HIGH LEVELS DURING AEROBIC GROWTH. NAPAB
CC       COMPLEX RECEIVES ELECTRONS FROM THE MEMBRANE-ANCHORED TETRAHEME
CC       NAPC PROTEIN, THUS ALLOWING ELECTRON FLOW BETWEEN MEMBRANE AND
CC       PERIPLASM. ESSENTIAL FUNCTION FOR NITRATE ASSIMILATION AND MAY
CC       HAVE A ROLE IN ANAEROBIC METABOLISM.
CC   -!- SUBUNIT: HETERODIMER OF A CATALYTIC SUBUNIT AND A CYTOCHROME C (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- INDUCTION: INDUCED BY NITRATE. NOT REPRESSED BY AMMONIUM OR
CC       OXYGEN.
CC   -!- PTM: BINDS TWO HEME GROUPS PER MOLECULE (BY SIMILARITY).
CC   -!- SIMILARITY: STRONG, TO OTHER BACTERIAL NAPB.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z46806; -; NOT_ANNOTATED_CDS.
DR   InterPro; IPR000345; CytC_heme_bind.
DR   PROSITE; PS00190; CYTOCHROME_C; 2.
KW   Electron transport; Heme; Periplasmic; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    154       DIHEME CYTOCHROME C NAPB.
FT   MOD_RES      25     25       PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).
FT   BINDING      82     82       HEME 1 (COVALENT) (BY SIMILARITY).
FT   BINDING      85     85       HEME 1 (COVALENT) (BY SIMILARITY).
FT   METAL        86     86       IRON 1 (HEME AXIAL LIGAND)
FT                                (BY SIMILARITY).
FT   BINDING     122    122       HEME 2 (COVALENT) (BY SIMILARITY).
FT   BINDING     125    125       HEME 2 (COVALENT) (BY SIMILARITY).
FT   METAL       126    126       IRON 2 (HEME AXIAL LIGAND)
FT                                (BY SIMILARITY).
SQ   SEQUENCE   154 AA;  16908 MW;  63CE7404A5864977 CRC64;

  Query Match              3.7%;  Score 6;  DB 1;  Length 154;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      24 PETITD 29
           ||||||
Db      46 PETITD 51


RESULT 23
LSPA_STACA
ID   LSPA_STACA     STANDARD;      PRT;   159 AA.
AC   Q59835;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Lipoprotein signal peptidase (EC 3.4.23.36) (Prolipoprotein signal
DE   peptidase) (Signal peptidase II) (SPase II).
GN   LSPA OR LSP.
OS   Staphylococcus carnosus.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1281;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TM300;
RX   MEDLINE=95247010; PubMed=7729667;
RA   Witke C., Goetz F.;
RT   "Cloning and nucleotide sequence of the signal peptidase II (lsp)-gene
RT   from Staphylococcus carnosus.";
RL   FEMS Microbiol. Lett. 126:233-240(1995).
CC   -!- FUNCTION: THIS PROTEIN SPECIFICALLY CATALYZES THE REMOVAL OF
CC       SIGNAL PEPTIDES FROM PROLIPOPROTEINS.
CC   -!- CATALYTIC ACTIVITY: Cleavage of N-terminal leader sequences from
CC       membrane prolipoproteins. Hydrolyzes Xaa-Xbb-Xcc-|-Cys, in which
CC       Xaa is hydrophobic (preferably Leu), Xbb is often Ser or Ala, Xcc
CC       is often Gly or Ala, and the Cys is alkylated on sulfur with a
CC       diacylglyceryl group.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A8.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78084; CAA54991.1; -.
DR   MEROPS; A08.001; -.
DR   InterPro; IPR001872; SigPTase_A8.
DR   Pfam; PF01252; Peptidase_A8; 1.
```

```
DR   PRINTS; PR00781; LIPOSIGPTASE.
DR   ProDom; PD004304; SigPTase_A8; 1.
DR   TIGRFAMs; TIGR00077; lspA; 1.
DR   PROSITE; PS00855; SPASE_II; 1.
KW   Hydrolase; Aspartyl protease; Transmembrane.
FT   DOMAIN        1      3       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      4     24       POTENTIAL.
FT   DOMAIN       25     63       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     64     84       POTENTIAL.
FT   DOMAIN       85     87       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     88    108       POTENTIAL.
FT   DOMAIN      109    130       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    131    151       POTENTIAL.
FT   DOMAIN      152    159       CYTOPLASMIC (POTENTIAL).
FT   ACT_SITE    109    109       BY SIMILARITY.
FT   ACT_SITE    136    136       BY SIMILARITY.
SQ   SEQUENCE   159 AA;  17630 MW;  F9DF08E222811FBD CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 159;
  Best Local Similarity   100.0%;  Pred. No. 78;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    101 IAVLIL 106
          ||||||
Db     15 IAVLIL 20


RESULT 24
RL6_BACSU
ID   RL6_BACSU      STANDARD;      PRT;   178 AA.
AC   P46898;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   50S ribosomal protein L6 (BL10).
GN   RPLF.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / Marburg;
RX   MEDLINE=96186897; PubMed=8635744;
RA   Suh J.W., Boylan S.A., Oh S.H., Price C.W.;
RT   "Genetic and transcriptional organization of the Bacillus subtilis
RT   spc-alpha region.";
RL   Gene 169:17-23(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124188; PubMed=8969501;
RA   Yasumoto K., Liu H., Jeong S.M., Ohashi Y., Kakinuma S.,
RA   Tanaka K., Kawamura F., Yoshikawa H., Takahashi H.;
RT   "Sequence analysis of a 50 kb region between spo0H and rrnH on the
RT   Bacillus subtilis chromosome.";
RL   Microbiology 142:3039-3046(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
CC       LOCATED AT THE AMINOACYL-TRNA BINDING SITE OF THE
CC       PEPTIDYLTRANSFERASE CENTER.
CC   -!- SIMILARITY: BELONGS TO THE L6P FAMILY OF RIBOSOMAL PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L47971; AAB06814.1; -.
DR   EMBL; D64125; BAA10981.1; -.
DR   EMBL; Z99104; CAB11907.1; -.
DR   HSSP; P02391; 1RL6.
DR   SubtiList; BG11408; rplF.
DR   InterPro; IPR000702; Ribosomal_L6.
DR   InterPro; IPR002358; Ribosomal_L6_1.
DR   Pfam; PF00347; Ribosomal_L6; 1.
DR   PRINTS; PR00059; RIBOSOMALL6.
DR   ProDom; PD002236; Ribosomal_L6; 1.
DR   PROSITE; PS00525; RIBOSOMAL_L6_1; 1.
KW   Ribosomal protein; rRNA-binding; Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   178 AA;  19378 MW;  A560F58CD68CFCDC CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 178;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    142 DVTVTL 147
          ||||||
Db     13 DVTVTL 18


RESULT 25
PTH_CHLMU
ID   PTH_CHLMU      STANDARD;      PRT;   179 AA.
AC   P49607;
DT   01-FEB-1996 (Rel. 33, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Peptidyl-tRNA hydrolase (EC 3.1.1.29) (PTH).
GN   PTH OR TC0183.
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MoPn;
RX   MEDLINE=96186910; PubMed=8635758;
RA   de la Vega F.M., Galindo J.M., Old I.G., Guarneros G.;
RT   "Microbial genes homologous to the peptidyl-tRNA hydrolase-encoding
```

```
RT   gene of Escherichia coli.";
RL   Gene 169:97-100(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MoPn / Nigg;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K.,
RA   Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA   Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G.,
RA   Salzberg S.L., Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE OF 27-194 FROM N.A.
RC   STRAIN=MoPn;
RX   MEDLINE=91072247; PubMed=2254267;
RA   Engel J.N., Pollack J., Perara E., Ganem D.;
RT   "Heat shock response of murine Chlamydia trachomatis.";
RL   J. Bacteriol. 172:6959-6972(1990).
CC   -!- FUNCTION: THE NATURAL SUBSTRATE FOR THIS ENZYME MAY BE PEPTIDYL-
CC       TRNAS WHICH DROP OFF THE RIBOSOME DURING PROTEIN SYNTHESIS
CC       (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N-substituted aminoacyl-tRNA + H(2)O = N-
CC       substituted amino acid + tRNA.
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE PTH FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U31570; AAB06184.1; -.
DR   EMBL; AE002285; AAF39057.1; -.
DR   EMBL; M62820; AAA23133.1; -.
DR   PIR; C37840; C37840.
DR   HSSP; P23932; 2PTH.
DR   TIGR; TC0183; -.
DR   InterPro; IPR001328; Pept_tRNA_hydro.
DR   Pfam; PF01195; Pept_tRNA_hydro; 1.
DR   ProDom; PD005324; Pept_tRNA_hydro; 1.
DR   TIGRFAMs; TIGR00447; pth; 1.
DR   PROSITE; PS01195; PEPT_TRNA_HYDROL_1; 1.
DR   PROSITE; PS01196; PEPT_TRNA_HYDROL_2; 1.
KW   Hydrolase; Complete proteome.
FT   CONFLICT    146    146       V -> G (IN REF. 1 AND 3).
FT   CONFLICT    161    161       S -> G (IN REF. 1 AND 3).
FT   CONFLICT    178    179       GF -> AFNLKGIYCSLFEKKSS (IN REF. 1 AND
FT                                3).
SQ   SEQUENCE   179 AA;  19915 MW;  5D682B5269C0564B CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 179;
  Best Local Similarity   100.0%;  Pred. No. 86;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 VLSNFS 47
           ||||||
Db     150 VLSNFS 155


RESULT 26
RUVA_FUSNN
ID   RUVA_FUSNN      STANDARD;      PRT;   194 AA.
AC   Q8REJ7;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Holliday junction DNA helicase ruvA.
GN   RUVA OR FN1104.
OS   Fusobacterium nucleatum (subsp. nucleatum).
OC   Bacteria; Fusobacteria; Fusobacterium.
OX   NCBI_TaxID=76856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25586;
RX   MEDLINE=21886394; PubMed=11889109;
RA   Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA   Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA   Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA   Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA   Fonstein M., Kyrpides N., Overbeek R.;
RT   "Genome sequence and analysis of the oral bacterium Fusobacterium
RT   nucleatum strain ATCC 25586.";
RL   J. Bacteriol. 184:2005-2018(2002).
CC   -!- FUNCTION: The ruvA-ruvB complex in the presence of ATP renatures
CC       cruciform structure in supercoiled DNA with palindromic sequence,
CC       indicating that it may promote strand exchange reactions in
CC       homologous recombination. RuvAB is an helicase that mediates the
CC       Holliday junction migration by localized denaturation and
CC       reanneling. RuvA stimulates, in the presence of DNA, the weak
CC       ATPase activity of ruvB (By similarity).
CC   -!- SUBUNIT: Forms a complex with ruvB.
CC   -!- SIMILARITY: BELONGS TO THE RUVA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE010615; AAL95300.1; -.
KW   DNA repair; SOS response; DNA-binding; DNA recombination; Helicase;
KW   Complete proteome.
SQ   SEQUENCE   194 AA;  22579 MW;  D40381C5BFB4B0F3 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 194;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 VLSNFS 47
           ||||||
Db      88 VLSNFS 93


RESULT 27
RL3_BUCAI
ID   RL3_BUCAI      STANDARD;      PRT;   209 AA.
AC   P57591;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   50S ribosomal protein L3.
GN   RPLC OR BU524.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tokyo 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND MAY
CC       PARTICIPATE IN THE FORMATION OF THE PEPTIDYLTRANSFERASE CENTER OF
CC       THE RIBOSOME (BY SIMILARITY).
```

```
CC   -!- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP001119; BAB13217.1; -.
DR   InterPro; IPR000597; Ribosomal_L3.
DR   Pfam; PF00297; Ribosomal_L3; 1.
DR   ProDom; PD001374; Ribosomal_L3; 1.
DR   PROSITE; PS00474; RIBOSOMAL_L3; 1.
KW   Ribosomal protein; rRNA-binding; Complete proteome.
SQ   SEQUENCE   209 AA;  22841 MW;  CF0D7F60C828C76F CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 209;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      46 FSGTVK 51
           ||||||
Db     118 FSGTVK 123


RESULT 28
MUC3_HUMAN
ID   MUC3_HUMAN     STANDARD;      PRT;   213 AA.
AC   Q02505; Q02506;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Mucin 3A (Intestinal mucin 3A) (Fragments).
GN   MUC3A OR MUC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Small intestine;
RX   MEDLINE=90365738; PubMed=2393399;
RA   Gum J.R. Jr., Hicks J.W., Swallow D.M., Lagace R.L., Byrd J.C.,
RA   Lamport D.T.A., Siddiki B., Kim Y.S.;
RT   "Molecular cloning of cDNAs derived from a novel human intestinal
RT   mucin gene.";
RL   Biochem. Biophys. Res. Commun. 171:407-415(1990).
CC   -!- FUNCTION: MAJOR GLYCOPROTEIN COMPONENT OF A VARIETY OF MUCUS GELS.
CC       THOUGHT TO PROVIDE A PROTECTIVE, LUBRICATING BARRIER AGAINST
CC       PARTICLES AND INFECTIOUS AGENTS AT MUCOSAL SURFACES.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: BROAD SPECIFICITY; SMALL INTESTINE, COLON AND
CC       COLONIC TUMORS.
CC   -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M55406; AAA63773.1; -.
DR   EMBL; M55405; AAA63772.1; -.
DR   PIR; A35690; A35690.
DR   PIR; B35690; B35690.
DR   Genew; HGNC:7513; MUC3A.
DR   MIM; 158371; -.
KW   Glycoprotein; Repeat.
FT   NON_TER       1      1
FT   DOMAIN        <1   >213       REPEATS.
FT   REPEAT        <1     9
FT   REPEAT        10     26
FT   REPEAT        27     49
FT   REPEAT        50     66
FT   REPEAT        67     83
FT   REPEAT        84    100
FT   REPEAT       101    117
FT   REPEAT       118   >128
FT   NON_CONS     128    129
FT   REPEAT      <129    145
FT   REPEAT       146    162
FT   REPEAT       163    179
FT   REPEAT       180    196
FT   REPEAT       197   >213
FT   NON_TER      213    213
SQ   SEQUENCE   213 AA;  22042 MW;  064EC77A237D786E CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 213;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      61 TTSETP 66
           ||||||
Db       3 TTSETP 8


RESULT 29
RPB5_KLUMA
ID   RPB5_KLUMA     STANDARD;      PRT;   215 AA.
AC   Q9P4B9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA-directed RNA polymerases II 24 kDa polypeptide (EC 2.7.7.6)
DE   (RNA polymerase II subunit 5).
GN   RPB5.
OS   Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX   NCBI_TaxID=4911;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12424;
RX   MEDLINE=20432119; PubMed=10974568;
RA   Ladriere J.-M., Georis I., Guerineau M., Vandenhaute J.;
RT   "Kluyveromyces marxianus exhibits an ancestral Saccharomyces
RT   cerevisiae genome organization downstream of ADH2.";
RL   Gene 255:83-91(2000).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 15
CC       DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS SHARED BY ALL 3 YEAST RNA
CC       POLYMERASES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC       FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC       PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC       III FOR 5S AND TRNA GENES.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEBACTERIA RPOH / EUKARYOTIC RPB5
CC       RNA POLYMERASE SUBUNIT FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
```

```
DR   EMBL; AF225206; AAF91238.1; -.
DR   HSSP; P20434; 1DZF.
DR   InterPro; IPR000783; RNA_pol_H_23kD.
DR   Pfam; PF01191; RNA_pol_H; 1.
DR   ProDom; PD005155; RNA_pol_H_23kD; 1.
DR   PROSITE; PS01110; RNA_POL_H_23KD; 1.
KW   Transferase; DNA-directed RNA polymerase; Transcription;
KW   Nuclear protein.
SQ   SEQUENCE   215 AA;  24954 MW;  10D7EA1754B3ACF0 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 PVALYL 84
           ||||||
Db     183 PVALYL 188


RESULT 30
RPB5_YEAST
ID   RPB5_YEAST     STANDARD;      PRT;   215 AA.
AC   P20434; Q02121;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide
DE   (EC 2.7.7.6) (ABC27).
GN   RPB5 OR YBR154C OR YBR1204.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90249736; PubMed=2186966;
RA   Woychik N.A., Liao S.-M., Kolodziej P.A., Young R.A.;
RT   "Subunits shared by eukaryotic nuclear RNA polymerases.";
RL   Genes Dev. 4:313-323(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=93255907; PubMed=8488729;
RA   Baur A., Schaaff-Gerstenschlaeger I., Boles E., Miosga T., Rose M.,
RA   Zimmermann F.K.;
RT   "Sequence of a 4.8 kb fragment of Saccharomyces cerevisiae chromosome
RT   II including three essential open reading frames.";
RL   Yeast 9:289-293(1993).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=20300896; PubMed=10841537;
RA   Todone F., Weinzierl R.O.J., Brick P., Onesti S.;
RT   "Crystal structure of RPB5, a universal eukaryotic RNA polymerase
RT   subunit and transcription factor interaction target.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:6306-6310(2000).
CC   -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC       OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC       SUBSTRATES.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM 9 TO 15
CC       DIFFERENT POLYPEPTIDES. THIS SUBUNIT IS SHARED BY ALL 3 YEAST RNA
CC       POLYMERASES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC       FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC       PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC       III FOR 5S AND TRNA GENES.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEBACTERIA RPOH / EUKARYOTIC RPB5
CC       RNA POLYMERASE SUBUNIT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53287; CAA37381.1; -.
DR   EMBL; X71329; CAA50472.1; -.
DR   EMBL; S59774; AAC60556.1; -.
DR   EMBL; Z36023; CAA85113.1; -.
DR   PIR; A34588; A34588.
DR   PDB; 1DZF; 09-JUN-00.
DR   SGD; S0000358; RPB5.
DR   InterPro; IPR000783; RNA_pol_H_23kD.
DR   Pfam; PF01191; RNA_pol_H; 1.
DR   ProDom; PD005155; RNA_pol_H_23kD; 1.
DR   PROSITE; PS01110; RNA_POL_H_23KD; 1.
KW   Transferase; DNA-directed RNA polymerase; Transcription;
KW   Nuclear protein; 3D-structure.
FT   CONFLICT     37     37       L -> W (IN REF. 2; AAC60556).
SQ   SEQUENCE   215 AA;  25079 MW;  2A15F7114D69D829 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      79 PVALYL 84
           ||||||
Db     183 PVALYL 188


RESULT 31
ALKB_SALTY
ID   ALKB_SALTY     STANDARD;      PRT;   216 AA.
AC   P37462;
DT   01-OCT-1994 (Rel. 30, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Alkylated DNA repair protein alkB.
GN   ALKB OR STM2264.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
RN   [2]
RP   SEQUENCE OF 1-24 FROM N.A.
RX   MEDLINE=91267928; PubMed=1904855;
RA   Hakura A., Morimoto K., Sofuni T., Nohmi T.;
RT   "Cloning and characterization of the Salmonella typhimurium ada gene,
RT   which encodes O6-methylguanine-DNA methyltransferase.";
RL   J. Bacteriol. 173:3663-3672(1991).
CC   -!- FUNCTION: PROBABLY INVOLVED IN THE REPAIR OF ALKYLATED DNA.
CC   -!- SIMILARITY: BELONGS TO THE ALKB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; AE008801; AAL21166.1; -.
DR   EMBL; D90221; -; NOT_ANNOTATED_CDS.
DR   PIR; B39433; B39433.
DR   StyGene; SG10006; alkB.
KW   DNA repair; Complete proteome.
FT   CONFLICT     19     19       A -> R (IN REF. 2).
SQ   SEQUENCE   216 AA;  23787 MW;  8F320C7B9CABFAB6 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 216;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     74 TDKPWP 79
          ||||||
Db     85 TDKPWP 90


RESULT 32
CAT_PROMI
ID   CAT_PROMI      STANDARD;      PRT;   217 AA.
AC   P07641;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   Chloramphenicol acetyltransferase (EC 2.3.1.28).
GN   CAT.
OS   Proteus mirabilis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Proteus.
OX   NCBI_TaxID=584;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM13;
RX   MEDLINE=86008040; PubMed=3900035;
RA   Charles I.G., Keyte J.W., Shaw W.V.;
RT   "Nucleotide sequence analysis of the cat gene of Proteus mirabilis:
RT   comparison with the type I (Tn9) cat gene.";
RL   J. Bacteriol. 164:123-129(1985).
CC   -!- FUNCTION: THIS ENZYME IS AN EFFECTOR OF CHLORAMPHENICOL RESISTANCE
CC       IN BACTERIA.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + chloramphenicol = CoA +
CC       chloramphenicol 3-acetate.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- SIMILARITY: BELONGS TO THE CHLORAMPHENICOL ACETYLTRANSFERASE
CC       FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M11587; AAA25655.1; -.
DR   PIR; A24651; A24651.
DR   HSSP; P00484; 1QCA.
DR   InterPro; IPR001707; CAT.
DR   Pfam; PF00302; CAT; 1.
DR   ProDom; PD002660; CAT; 1.
DR   PROSITE; PS00100; CAT; 1.
KW   Antibiotic resistance; Transferase; Acyltransferase.
FT   ACT_SITE    193    193       BY SIMILARITY.
SQ   SEQUENCE   217 AA;  25313 MW;  08ABB443F9FC41C2 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 217;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 LVVDLS 14
          ||||||
Db      9 LVVDLS 14


RESULT 33
R10A_SPOFR
ID   R10A_SPOFR     STANDARD;      PRT;   217 AA.
AC   Q963B6;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   60S ribosomal protein L10a.
GN   RPL10A.
OS   Spodoptera frugiperda (Fall armyworm).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC   Ditrysia; Noctuoidea; Noctuidae; Amphipyrinae; Spodoptera.
OX   NCBI_TaxID=7108;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Landais I., Ogliastro M., Mita K., Nohata J., Lopez-Ferber M.,
RA   Duonor-Cerutti M., Fournier P., Devauchelle G.;
RT   "Full-length ribosomal protein sequence from an EST library of
RT   Spodoptera frugiperda cells (Sf9).";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE L1P FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF391092; AAK76990.1; -.
DR   InterPro; IPR002143; Ribosomal_L1.
DR   Pfam; PF00687; Ribosomal_L1; 1.
DR   PROSITE; PS01199; RIBOSOMAL_L1; FALSE_NEG.
KW   Ribosomal protein.
SQ   SEQUENCE   217 AA;  24518 MW;  9E284AAE1FFA082C CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 217;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     46 FSGTVK 51
          ||||||
Db     49 FSGTVK 54


RESULT 34
VTI1_YEAST
ID   VTI1_YEAST     STANDARD;      PRT;   217 AA.
AC   Q04338;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Vesicle transport v-SNARE protein VTI1.
GN   VTI1 OR YMR197C OR YM9646.10C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97344273; PubMed=9199167;
RA   Fischer von Mollard G., Nothwehr S.F., Stevens T.H.;
RT   "The yeast v-SNARE Vti1p mediates two vesicle transport pathways
RT   through interactions with the t-SNAREs Sed5p and Pep12p.";
RL   J. Cell Biol. 137:1511-1524(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
```

```
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=98062344; PubMed=9398683;
RA   Lupashin V.V., Pokrovskaya I.D., McNew J.A., Waters M.G.;
RT   "Characterization of a novel yeast SNARE protein implicated in Golgi
RT   retrograde traffic.";
RL   Mol. Biol. Cell 8:2659-2676(1997).
RN   [4]
RP   BINDING TO T-SNARES TLG1 AND TLG2.
RX   MEDLINE=98090455; PubMed=9427746;
RA   Holthuis J.C., Nichols B.J., Dhruvakumar S., Pelham H.R.;
RT   "Two syntaxin homologues in the TGN/endosomal system of yeast.";
RL   EMBO J. 17:113-126(1998).
RN   [5]
RP   INVOLVEMENT IN HOMOTYPIC VACUOLAR FUSION.
RX   MEDLINE=99315912; PubMed=10385523;
RA   Ungermann C., Fischer von Mollard G., Jensen O.N., Margolis N.,
RA   Stevens T.H., Wickner W.;
RT   "Three v-SNAREs and two t-SNAREs, present in a pentameric cis-SNARE
RT   complex on isolated vacuoles, are essential for homotypic fusion.";
RL   J. Cell Biol. 145:1435-1442(1999).
RN   [6]
RP   INVOLVEMENT IN MULTIPLE TRAFFICKING STEPS.
RX   MEDLINE=99287719; PubMed=10359592;
RA   Fischer von Mollard G., Stevens T.H.;
RT   "The Saccharomyces cerevisiae v-SNARE Vti1p is required for multiple
RT   membrane transport pathways to the vacuole.";
RL   Mol. Biol. Cell 10:1719-1732(1999).
CC   -!- FUNCTION: V-SNARE ESSENTIAL FOR YEAST CELL VIABILITY WHICH
CC       MEDIATES MULTIPLE VESICLE TRANSPORT PATHWAYS THROUGH INTERACTIONS
CC       WITH T-SNARES ON THE TARGET MEMBRANE OF DIFFERENT ORGANELLES.
CC       THESE INTERACTIONS ARE PROPOSED TO MEDIATE ASPECTS OF THE
CC       SPECIFICITY OF VESICLE TRAFFICKING AND TO PROMOTE FUSION OF THE
CC       LIPID BILAYERS.
CC   -!- SUBUNIT: FORMS A SNARE COMPLEX WITH THE CIS-GOLGI T-SNARE SED5 AND
CC       V-SNARES SFT1 AND YTK6, WHICH IS INVOLVED IN RETROGRADE TRAFFIC TO
CC       THE CIS-GOLGI COMPARTMENT. INTERACTS WITH THE PREVACUOLAR T-SNARE
CC       PEP12 IN TRAFFIC FROM THE TRANS-GOLGI NETWORK (TGN) TO THE
CC       PREVACUOLAR COMPARTMENT (PVC). INTERACTS WITH THE VACUOLAR T-SNARE
CC       VAM3 IN BIOSYNTHETIC TRANSPORT PATHWAYS TO THE VACUOLE AND IN
CC       HOMOTYPIC VACUOLAR FUSION. BINDS ALSO TO THE EARLY ENDOSOME T-
CC       SNARE TLG1 AND THE TGN T-SNARE TLG2.
CC   -!- SUBCELLULAR LOCATION: A SMALL PORTION IS LOCALIZED IN THE GOLGI
CC       APPARATUS, THE MAJORITY IS LOCALIZED IN THE PVC.
CC   -!- MISCELLANEOUS: TEMPERATURE-SENSITIVE (TS) MUTANTS VTI1-1 AND VTI1-
CC       2 EXHIBIT DEFECTS IN TGN TO PVC TRANSPORT AT THE NONPERMISSIVE
CC       TEMPERATURE. VTI1-11 MUTANT DISPLAY A BLOCK IN TRAFFIC TO THE PVC
CC       AND AN ADDITIONAL DEFECT IN RETROGRADE TRAFFIC TO THE CIS-GOLGI.
CC       VTI1-12 MUTANT HAS A CONSTITUTIVE BLOCK OF TRAFFIC FROM THE LATE
CC       GOLGI TO THE VACUOLE AND A TS BLOCK TO THE CIS-GOLGI COMPARTMENT.
CC   -!- SIMILARITY: BELONGS TO THE VTI1 FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 T-SNARE COILED-COIL HOMOLOGY DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF006074; AAC49745.1; -.
DR   EMBL; Z47815; CAA87819.1; -.
DR   SGD; S0004810; VTI1.
DR   InterPro; IPR000727; T_SNARE.
DR   SMART; SM00397; t_SNARE; 1.
DR   PROSITE; PS50192; T_SNARE; 1.
KW   Transmembrane; Transport; Protein transport; Coiled coil.
FT   DOMAIN      124     186       T-SNARE COILED-COIL HOMOLOGY.
FT   TRANSMEM    195     215       POTENTIAL.
FT   VARIANT       8       8       Y -> R (IN MUTANT VTI1-11).
FT   VARIANT      20      20       K -> R (IN MUTANT VTI1-11).
FT   VARIANT      40      40       H -> R (IN MUTANT VTI1-11).
FT   VARIANT      61      61       N -> S (IN MUTANT VTI1-11).
FT   VARIANT      73      73       K -> R (IN MUTANT VTI1-11).
FT   VARIANT      84      84       Q -> R (IN MUTANT VTI1-11).
FT   VARIANT     130     130       S -> P (IN MUTANT VTI1-2).
FT   VARIANT     141     141       A -> S (IN MUTANT VTI1-12).
FT   VARIANT     145     145       E -> G (IN MUTANT VTI1-11).
FT   VARIANT     145     145       E -> K (IN MUTANT VTI1-1).
FT   VARIANT     148     148       G -> R (IN MUTANT VTI1-1).
FT   VARIANT     155     155       L -> F (IN MUTANT VTI1-11).
FT   VARIANT     158     158       Q -> R (IN MUTANT VTI1-12).
SQ   SEQUENCE   217 AA;  24668 MW;  E55B4DA3B40FAD67 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 217;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     101 IAVLIL 106
           ||||||
Db     201 IAVLIL 206


RESULT 35
GCH1_HAEIN
ID   GCH1_HAEIN     STANDARD;      PRT;   218 AA.
AC   P43866;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   GTP cyclohydrolase I (EC 3.5.4.16) (GTP-CH-I).
GN   FOLE OR HI1447.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: GTP + 2 H(2)O = formate + 2-amino-4-hydroxy-6-
CC       (erythro-1,2,3-trihydroxypropyl)dihydropteridine triphosphate.
CC   -!- PATHWAY: Tetrahydrofolate biosynthesis; first step.
CC   -!- SUBUNIT: HOMOPOLYMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32823; AAC23097.1; -.
DR   HSSP; P27511; 1A8R.
DR   TIGR; HI1447; -.
DR   InterPro; IPR001474; GTP_cyclohydroI.
DR   Pfam; PF01227; GTP_cyclohydroI; 1.
DR   ProDom; PD003330; GTP_cyclohydroI; 1.
DR   TIGRFAMs; TIGR00063; folE; 1.
DR   PROSITE; PS00859; GTP_CYCLOHYDROL_1_1; 1.
```

```
DR   PROSITE; PS00860; GTP_CYCLOHYDROL_1_2; 1.
KW   One-carbon metabolism; Hydrolase; Complete proteome.
FT   DISULFID    109    180       BY SIMILARITY.
SQ   SEQUENCE   218 AA;  24945 MW;  BCC60F0038158D38 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 218;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 SAYGGV 42
           ||||||
Db     196 SAYGGV 201


RESULT 36
GLNP_ECOLI
ID   GLNP_ECOLI     STANDARD;      PRT;   219 AA.
AC   P10345; P76825;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glutamine transport system permease protein glnP.
GN   GLNP OR B0810 OR Z1032 OR ECS0888.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=87115160; PubMed=3027504;
RA   Nohno T., Saito T., Hong J.;
RT   "Cloning and complete nucleotide sequence of the Escherichia coli
RT   glutamine permease operon (glnHPQ).";
RL   Mol. Gen. Genet. 205:260-269(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC       FOR GLUTAMINE; PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF THE
CC       SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- INDUCTION: BY LACK OF GLUTAMINE.
CC   -!- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC       SYSTEM PERMEASE FAMILY. HISMQ SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X14180; CAA32383.1; -.
DR   EMBL; AE000183; AAC73897.1; -.
DR   EMBL; D90718; BAA35482.1; -.
DR   EMBL; D90719; BAA35492.1; -.
DR   EMBL; AE005262; AAG55182.1; -.
DR   EMBL; AP002553; BAB34311.1; -.
DR   PIR; S03182; QRECGP.
DR   EcoGene; EG10388; glnP.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF00528; BPD_transp; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
KW   Transport; Amino-acid transport; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   TRANSMEM     23     43       POTENTIAL.
FT   TRANSMEM     54     74       POTENTIAL.
FT   TRANSMEM     89    109       POTENTIAL.
FT   TRANSMEM    151    171       POTENTIAL.
FT   TRANSMEM    188    208       POTENTIAL.
SQ   SEQUENCE   219 AA;  24364 MW;  20A94A31BE02121C CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 219;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      90 AGGLVI 95
           ||||||
Db      31 AGGLVI 36


RESULT 37
VV_PI2H
ID   VV_PI2H        STANDARD;      PRT;   225 AA.
AC   P19847;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Nonstructural protein V.
GN   P/V.
OS   Human parainfluenza 2 virus (PIV-2).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX   NCBI_TaxID=11212;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90281608; PubMed=2162111;
RA   Southern J.A., Precious B., Randall R.E.;
RT   "Two nontemplated nucleotide additions are required to generate the P
RT   mRNA of parainfluenza virus type 2 since the RNA genome encodes
```

```
RT   protein V.";
RL   Virology 177:388-390(1990).
CC   -!- ALTERNATIVE PRODUCTS: THE P PROTEIN IS COTERMINAL WITH THE V
CC       PROTEIN, AND IS ENCODED BY A SEPARATE MRNA WHICH DIFFERS BY TWO
CC       NONTEMPLATED NUCLEOTIDES. THE SEQUENCE OF P AND V ARE IDENTICAL
CC       UP TO RESIDUE 164.
CC   -!- SIMILARITY: TO THE V PROTEIN OF SIMIAN VIRUS 5.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37748; AAA46808.1; -.
DR   PIR; A35322; MNNZP2.
DR   InterPro; IPR004897; Paramyx_P_V.
DR   Pfam; PF03210; Paramyx_P_V; 1.
KW   Nonstructural protein; RNA editing.
FT   DOMAIN      193    221       CYS-RICH.
SQ   SEQUENCE   225 AA;  24121 MW;  3F96DA969F7F9CCC CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 PVSSAG 91
           ||||||
Db     148 PVSSAG 153


RESULT 38
VV_PI2HT
ID   VV_PI2HT       STANDARD;      PRT;   225 AA.
AC   P23057;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Nonstructural protein V.
GN   P/V.
OS   Human parainfluenza 2 virus (strain Toshiba) (PIV-2).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX   NCBI_TaxID=11214;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90281574; PubMed=2162103;
RA   Ohgimoto S., Bando H., Kawano M., Okamoto K., Kondo K., Tsurudome M.,
RA   Nishio M., Ito Y.;
RT   "Sequence analysis of P gene of human parainfluenza type 2 virus: P
RT   and cysteine-rich proteins are translated by two mRNAs that differ by
RT   two nontemplated G residues.";
RL   Virology 177:116-123(1990).
CC   -!- ALTERNATIVE PRODUCTS: THE P PROTEIN IS COTERMINAL WITH THE V
CC       PROTEIN, AND IS ENCODED BY A SEPARATE MRNA WHICH DIFFERS BY TWO
CC       NONTEMPLATED NUCLEOTIDES. THE SEQUENCE OF P AND V ARE IDENTICAL
CC       UP TO RESIDUE 164.
CC   -!- SIMILARITY: TO THE V PROTEIN OF SIMIAN VIRUS 5.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M37751; AAA46803.1; -.
DR   EMBL; X57559; CAA40784.1; -.
DR   PIR; A35313; MNNZVT.
DR   PIR; S16660; S16660.
DR   InterPro; IPR004897; Paramyx_P_V.
DR   Pfam; PF03210; Paramyx_P_V; 1.
KW   Nonstructural protein; RNA editing.
FT   DOMAIN      193    221       CYS-RICH.
SQ   SEQUENCE   225 AA;  24151 MW;  3F9D703794CF8DDD CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 PVSSAG 91
           ||||||
Db     148 PVSSAG 153


RESULT 39
TRYP_PIG
ID   TRYP_PIG       STANDARD;      PRT;   231 AA.
AC   P00761;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Trypsin precursor (EC 3.4.21.4).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE OF 1-10.
RA   Charles M., Rovery M., Guidoni A.A., Desnuelle P.;
RT   "On trypsinogen and trypsin of pig.";
RL   Biochim. Biophys. Acta 69:115-129(1963).
RN   [2]
RP   SEQUENCE OF 9-231.
RX   MEDLINE=73258692; PubMed=4738933;
RA   Hermodson M.A., Ericsson L.H., Neurath H., Walsh K.A.;
RT   "Determination of the amino acid sequence of porcine trypsin by
RT   sequenator analysis.";
RL   Biochemistry 12:3146-3153(1973).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX   MEDLINE=93187998; PubMed=8445634;
RA   Huang Q., Liu S., Tang Y.;
RT   "Refined 1.6-A resolution crystal structure of the complex formed
RT   between porcine beta-trypsin and MCTI-A, a trypsin inhibitor of the
RT   squash family. Detailed comparison with bovine beta-trypsin and its
RT   complex.";
RL   J. Mol. Biol. 229:1022-1030(1993).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RX   MEDLINE=92201369; PubMed=1551419;
RA   Huang Q., Liu S., Tang Y., Zeng F., Qian R.;
RT   "Amino acid sequencing of a trypsin inhibitor by refined 1.6 A X-ray
RT   crystal structure of its complex with porcine beta-trypsin.";
RL   FEBS Lett. 297:143-146(1992).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=95035057; PubMed=7947985;
RA   Huang Q., Wang Z., Li Y., Liu S., Tang Y.;
RT   "Refined 1.8-A resolution crystal structure of the porcine epsilon-
RT   trypsin.";
RL   Biochim. Biophys. Acta 1209:77-82(1994).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF COMPLEX WITH LDTI.
RX   MEDLINE=97390427; PubMed=9242660;
RA   Stubbs M.T., Morenweiser R., Sturzebecher J., Bauer M., Bode W.,
RA   Huber R., Piechottka G.P., Matschiner G., Sommerhoff C.P., Fritz H.,
RA   Auerswald E.A.;
RT   "The three-dimensional structure of recombinant leech-derived
RT   tryptase inhibitor in complex with trypsin. Implications for the
RT   structure of human mast cell tryptase and its inhibition.";
RL   J. Biol. Chem. 272:19931-19937(1997).
RN   [7]
```

```
RP   X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF COMPLEX WITH LDTI.
RX   MEDLINE=98046095; PubMed=9384562;
RA   di Marco S., Priestle J.P.;
RT   "Structure of the complex of leech-derived tryptase inhibitor (LDTI)
RT   with trypsin and modeling of the LDTI-tryptase system.";
RL   Structure 5:1465-1474(1997).
CC   -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR   PIR; A00947; TRPGTR.
DR   PDB; 1MCT; 31-JAN-94.
DR   PDB; 1AKS; 12-FEB-97.
DR   PDB; 1EPT; 07-FEB-95.
DR   PDB; 1TFX; 21-JAN-98.
DR   PDB; 1LDT; 20-MAY-98.
DR   PDB; 1AN1; 01-JUL-98.
DR   PDB; 1AVW; 18-NOV-98.
DR   PDB; 1AVX; 18-NOV-98.
DR   MEROPS; S01.151; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Hydrolase; Serine protease; Digestion; Pancreas; Zymogen;
KW   3D-structure.
FT   PROPEP        1      8       ACTIVATION PEPTIDE.
FT   CHAIN         9    231       TRYPSIN.
FT   ACT_SITE     48     48       CHARGE RELAY SYSTEM.
FT   ACT_SITE     92     92       CHARGE RELAY SYSTEM.
FT   ACT_SITE    185    185       CHARGE RELAY SYSTEM.
FT   DISULFID     15    145
FT   DISULFID     33     49
FT   DISULFID    117    218
FT   DISULFID    124    191
FT   DISULFID    156    170
FT   DISULFID    181    205
FT   SITE        179    179       REQUIRED FOR SPECIFICITY (BY SIMILARITY).
FT   VARIANT      20     20       I -> V.
FT   STRAND       10     10
FT   STRAND       13     14
FT   TURN         17     18
FT   TURN         21     22
FT   STRAND       23     27
FT   STRAND       31     39
FT   TURN         40     41
FT   STRAND       42     45
FT   HELIX        47     49
FT   STRAND       55     58
FT   STRAND       62     62
FT   TURN         63     64
FT   STRAND       71     80
FT   TURN         82     83
FT   TURN         86     88
FT   TURN         90     91
FT   STRAND       94     98
FT   STRAND      112    112
FT   TURN        120    121
FT   STRAND      123    128
FT   STRAND      142    142
FT   STRAND      144    150
FT   HELIX       153    159
FT   TURN        161    163
FT   TURN        166    167
FT   STRAND      168    171
FT   TURN        174    175
FT   STRAND      179    179
FT   TURN        182    183
FT   TURN        185    186
FT   STRAND      188    191
FT   TURN        192    193
FT   STRAND      194    202
FT   STRAND      207    207
FT   TURN        208    209
FT   STRAND      210    210
FT   STRAND      212    216
FT   HELIX       217    219
FT   HELIX       221    230
SQ   SEQUENCE   231 AA;  24409 MW;  A0A125CF7FC138C2 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 231;
  Best Local Similarity  100.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 SGSSYP 58
           ||||||
Db     135 SGSSYP 140


RESULT 40
SODM_MAIZE
ID   SODM_MAIZE     STANDARD;      PRT;   235 AA.
AC   P09233;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Superoxide dismutase [Mn] 3.1, mitochondrial precursor (EC 1.15.1.1).
GN   SODA.4 OR SOD3.1 OR SOD3.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. W64A;
RX   MEDLINE=89051020; PubMed=2461225;
RA   Redinbaugh M.G., Wadsworth G.T., Scandalios J.G.;
RT   "Isolation and characterization of a cDNA for mitochondrial manganese
RT   superoxide dismutase (SOD-3) of maize and its relation to other
RT   manganese superoxide dismutases.";
RL   Biochim. Biophys. Acta 951:61-70(1988).
CC   -!- FUNCTION: Destroys radicals which are normally produced within the
CC       cells and which are toxic to biological systems.
CC   -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC   -!- COFACTOR: Manganese.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC   -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12540; CAA31058.1; -.
DR   PIR; S03839; S03839.
DR   HSSP; P04179; 1ABM.
DR   MaizeDB; 47587; -.
DR   InterPro; IPR001189; SODismutase.
DR   Pfam; PF00081; sodfe; 1.
DR   Pfam; PF02777; sodfe_C; 1.
DR   ProDom; PD000475; SODismutase; 1.
DR   PROSITE; PS00088; SOD_MN; 1.
KW   Oxidoreductase; Manganese; Mitochondrion; Transit peptide;
KW   Multigene family.
FT   TRANSIT       1     31       MITOCHONDRION.
FT   CHAIN        32    235       SUPEROXIDE DISMUTASE [MN] 3.1.
FT   METAL        59     59       MANGANESE (BY SIMILARITY).
```

```
FT   METAL       107    107       MANGANESE (BY SIMILARITY).
FT   METAL       196    196       MANGANESE (BY SIMILARITY).
FT   METAL       200    200       MANGANESE (BY SIMILARITY).
SQ   SEQUENCE   235 AA;  25545 MW;  AD51BAD0F44FDE56 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     144 TVTLPD 149
           ||||||
Db      34 TVTLPD 39


RESULT 41
CS31_ECOLI
ID   CS31_ECOLI     STANDARD;      PRT;   241 AA.
AC   P15483;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Chaperone protein CS3-1 precursor.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PB176;
RX   MEDLINE=90158116; PubMed=2576094;
RA   Jalajakumari M.B., Thomas C.J., Halter R., Manning P.A.;
RT   "Genes for biosynthesis and assembly of CS3 pili of CFA/II
RT   enterotoxigenic Escherichia coli: novel regulation of pilus
RT   production by bypassing an amber codon.";
RL   Mol. Microbiol. 3:1685-1695(1989).
RN   [2]
RP   SEQUENCE OF 1-168 FROM N.A.
RX   MEDLINE=89032631; PubMed=2903130;
RA   Boylan M., Smyth C.J., Scott J.R.;
RT   "Nucleotide sequence of the gene encoding the major subunit of CS3
RT   fimbriae of enterotoxigenic Escherichia coli.";
RL   Infect. Immun. 56:3297-3300(1988).
CC   -!- FUNCTION: THIS PROTEIN IS ESSENTIAL FOR THE BIOGENESIS OF
CC       MATURE CS3 PILI.
CC   -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE PERIPLASMIC PILUS CHAPERONE FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16944; CAA34815.1; -.
DR   EMBL; M35657; -; NOT_ANNOTATED_CDS.
DR   PIR; S07899; S07899.
DR   HSSP; P31697; 1BF8.
DR   InterPro; IPR001829; Pili_chaperone.
DR   Pfam; PF00345; pili_assembly; 1.
DR   Pfam; PF02753; pili_assembly_C; 1.
DR   PRINTS; PR00969; CHAPERONPILI.
DR   ProDom; PD001447; Pili_chaperone; 1.
DR   PROSITE; PS00635; PILI_CHAPERONE; 1.
KW   Chaperone; Fimbria; Periplasmic; Signal; Immunoglobulin domain.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    241       CHAPERONE PROTEIN CS3-1.
FT   DISULFID    113    144       POTENTIAL.
FT   CONFLICT     73     73       N -> S (IN REF. 2).
SQ   SEQUENCE   241 AA;  26908 MW;  DBC9767D163FCFA0 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 241;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     122 NIYANN 127
           ||||||
Db      18 NIYANN 23


RESULT 42
SDGF_RAT
ID   SDGF_RAT       STANDARD;      PRT;   243 AA.
AC   P24338;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Schwannoma-derived growth factor precursor (SDGF) (Amphiregulin) (AR).
GN   SDGF.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=91043095; PubMed=2234093;
RA   Kimura H., Fischer W.H., Schubert D.;
RT   "Structure, expression and function of a schwannoma-derived growth
RT   factor.";
RL   Nature 348:257-260(1990).
CC   -!- FUNCTION: AUTOCRINE GROWTH FACTOR AS WELL AS A MITOGEN FOR A BROAD
CC       RANGE OF TARGET CELLS, AMONG WHICH ASTROCYTES, SCHWANN CELLS AND
CC       FIBROBLASTS.
CC   -!- SIMILARITY: HIGH, TO AMPHIREGULIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55183; CAA38967.1; -.
DR   PIR; S13296; S13296.
DR   HSSP; Q99075; 1XDT.
DR   InterPro; IPR000561; EGF-like.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS00022; EGF_1; 1.
KW   Glycoprotein; Cytokine; Growth factor; EGF-like domain; Signal;
KW   Transmembrane.
FT   SIGNAL        1     24       POTENTIAL.
FT   PROPEP       25     96
FT   CHAIN        97    243       SCHWANNOMA-DERIVED GROWTH FACTOR.
FT   DOMAIN      133    173       EGF-LIKE.
FT   TRANSMEM    190    213       POTENTIAL.
FT   DISULFID    137    150       BY SIMILARITY.
FT   DISULFID    145    161       BY SIMILARITY.
FT   DISULFID    163    172       BY SIMILARITY.
FT   CARBOHYD     30     30       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    103    103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    236    236       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   243 AA;  26633 MW;  F59A8173D4921599 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLV 94
           ||||||
Db      45 SAGGLV 50
```

```
RESULT 43
VNST_PTPV
ID   VNST_PTPV      STANDARD;      PRT;   250 AA.
AC   P03516;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nonstructural protein NS-S.
OS   Punta toro phlebovirus.
OC   Viruses; ssRNA negative-strand viruses; Bunyaviridae; Phlebovirus.
OX   NCBI_TaxID=11587;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84276006; PubMed=6087547;
RA   Ihara T., Akashi H., Bishop D.H.L.;
RT   "Novel coding strategy (ambisense genomic RNA) revealed by sequence
RT   analyses of Punta toro phlebovirus S RNA.";
RL   Virology 136:293-306(1984).
CC   -!- MISCELLANEOUS: THIS PROTEIN MAY BE A TRANSCRIPTASE COMPONENT.
CC   -!- SIMILARITY: NS-S FROM PUNTA TORO, RIFT VALLEY FEVER, SANDFLY FEVER
CC       SICILIAN, TOSCANA, AND UUKUNIEMI VIRUSES ARE EVOLUTIONARY RELATED.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02736; AAA47115.1; -.
DR   PIR; A04108; MNVUPT.
KW   Nonstructural protein; Transcription.
SQ   SEQUENCE   250 AA;  29097 MW;  2C8909A1EDAD90D7 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      92 GLVIKA 97
           ||||||
Db      80 GLVIKA 85


RESULT 44
YD37_MYCLE
ID   YD37_MYCLE     STANDARD;      PRT;   251 AA.
AC   P53426; Q9CC73;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein ML1171.
GN   ML1171 OR B1549_C3_240.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Smith D.R., Robison K.;
RL   Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV1337.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00014; AAA50900.1; -.
DR   EMBL; AL583921; CAC31552.1; ALT_INIT.
DR   Leproma; ML1171; -.
DR   InterPro; IPR002610; Rhomboid.
DR   Pfam; PF01694; Rhomboid; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     48     68       POTENTIAL.
FT   TRANSMEM     88    108       POTENTIAL.
FT   TRANSMEM    110    130       POTENTIAL.
FT   TRANSMEM    132    152       POTENTIAL.
FT   TRANSMEM    158    178       POTENTIAL.
FT   TRANSMEM    184    204       POTENTIAL.
FT   TRANSMEM    209    229       POTENTIAL.
SQ   SEQUENCE   251 AA;  27064 MW;  BF847E050599DFC4 CRC64;

  Query Match             3.7%;  Score 6;  DB 1;  Length 251;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 AYGGVL 43
           ||||||
Db     192 AYGGVL 197


RESULT 45
DSBG_PSEAE
ID   DSBG_PSEAE     STANDARD;      PRT;   256 AA.
AC   Q9I106;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thiol:disulfide interchange protein dsbG precursor.
GN   DSBG OR PA2476.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: Involved in disulfide bond formation. Functions probably
CC       as a disulfide isomerase with a narrower substrate specificity
CC       than dsbC. DsbG is maintained in a reduced state by dsbD (By
CC       similarity).
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY. DSBC SUBFAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE004675; AAG05864.1; -.
DR   InterPro; IPR000063; Thiored.
DR   PROSITE; PS00194; THIOREDOXIN; FALSE_NEG.
KW   Redox-active center; Periplasmic; Signal; Complete proteome.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    256       THIOL:DISULFIDE INTERCHANGE PROTEIN DSBG.
FT   DISULFID    134    137       REDOX-ACTIVE (BY SIMILARITY).
SQ   SEQUENCE   256 AA;  28053 MW;  C1EB26268781CE23 CRC64;

  Query Match              3.7%;  Score 6;  DB 1;  Length 256;
  Best Local Similarity  100.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      81 ALYLTP 86
           ||||||
Db      65 ALYLTP 70


Search completed: November 28, 2002, 19:03:43
Job time : 35 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 18:59:56 ; Search time 81 Seconds
(without alignments)
409.550 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 161
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	92	57.1	166	2	Q9F5R9	Q9f5r9 escherichia
2	92	57.1	300	2	Q9S497	Q9s497 escherichia
3	92	57.1	300	2	Q9S496	Q9s496 escherichia
4	92	57.1	300	2	Q9F5S0	Q9f5s0 escherichia
5	90	55.9	300	16	Q8XBA6	Q8xba6 escherichia
6	73	45.3	300	2	Q9S495	Q9s495 escherichia
7	59	36.6	300	2	Q9S494	Q9s494 escherichia
8	59	36.6	300	2	Q9F5S1	Q9f5s1 escherichia
9	59	36.6	300	2	Q9R5Y2	Q9r5y2 escherichia
10	56	34.8	300	2	Q9S6R1	Q9s6r1 escherichia
11	56	34.8	300	2	O87634	O87634 escherichia
12	56	34.8	300	2	Q9F6Z7	Q9f6z7 escherichia
13	56	34.8	303	2	Q9AP05	Q9ap05 escherichia
14	50	31.1	129	2	Q9XBV8	Q9xbv8 escherichia
15	42	26.1	243	2	Q08858	Q08858 klebsiella
16	9	5.6	295	12	Q998F5	Q998f5 porcine cyt
17	9	5.6	304	16	Q8XAX2	Q8xax2 escherichia
18	9	5.6	550	13	Q9PT94	Q9pt94 brachydanio
19	9	5.6	550	13	Q9DG30	Q9dg30 brachydanio
20	8	5.0	539	16	Q9RWQ4	Q9rwq4 deinococcus
21	8	5.0	564	4	Q9H1Y7	Q9h1y7 homo sapien
22	8	5.0	570	4	Q8WWN2	Q8wwn2 homo sapien
23	8	5.0	599	12	Q80929	Q80929 human papil
24	8	5.0	860	10	Q9SHB7	Q9shb7 arabidopsis
25	8	5.0	1036	4	Q8WWN1	Q8wwn1 homo sapien
26	8	5.0	1239	5	O01505	O01505 caenorhabdi
27	8	5.0	1873	2	Q9Z4N7	Q9z4n7 enterococcu
28	7	4.3	63	5	Q9BJG6	Q9bjg6 plasmodium
29	7	4.3	73	4	O75702	O75702 homo sapien
30	7	4.3	74	5	Q9ND29	Q9nd29 plasmodium
31	7	4.3	87	16	O31846	O31846 bacillus su
32	7	4.3	107	5	Q9N3B4	Q9n3b4 caenorhabdi
33	7	4.3	119	10	Q9FE65	Q9fe65 arabidopsis
34	7	4.3	120	10	Q9LJW6	Q9ljw6 arabidopsis
35	7	4.3	128	16	Q9RXH5	Q9rxh5 deinococcus
36	7	4.3	162	16	Q8U8J1	Q8u8j1 agrobacteri
37	7	4.3	177	2	Q8VV80	Q8vv80 colwellia m
38	7	4.3	179	5	Q9U7J9	Q9u7j9 plasmodium
39	7	4.3	188	16	Q9PDQ1	Q9pdq1 xylella fas
40	7	4.3	210	16	Q9CHD6	Q9chd6 lactococcus
41	7	4.3	248	16	O05437	O05437 mycobacteri
42	7	4.3	248	16	Q8VIR0	Q8vir0 mycobacteri
43	7	4.3	251	16	Q50195	Q50195 mycobacteri
44	7	4.3	256	16	Q8R805	Q8r805 thermoanaer
45	7	4.3	283	10	Q9ZVF6	Q9zvf6 arabidopsis
46	7	4.3	296	10	Q9LS70	Q9ls70 arabidopsis
47	7	4.3	298	16	Q97IP7	Q97ip7 clostridium
48	7	4.3	304	17	Q9UZU9	Q9uzu9 pyrococcus
49	7	4.3	311	5	O44789	O44789 caenorhabdi
50	7	4.3	315	16	Q98L41	Q98l41 rhizobium l
51	7	4.3	328	10	Q41401	Q41401 sesbania ro
52	7	4.3	366	5	Q9N680	Q9n680 drosophila
53	7	4.3	383	16	Q8ZG16	Q8zg16 yersinia pe
54	7	4.3	396	2	O06037	O06037 lactococcus
55	7	4.3	397	5	Q9V5I8	Q9v5i8 drosophila
56	7	4.3	414	16	Q8UGJ7	Q8ugj7 agrobacteri
57	7	4.3	416	10	Q9LT16	Q9lt16 arabidopsis
58	7	4.3	417	2	Q9RPA9	Q9rpa9 streptomyce
59	7	4.3	443	16	Q9CEP2	Q9cep2 lactococcus
60	7	4.3	457	16	Q9CPF2	Q9cpf2 pasteurella
61	7	4.3	472	10	Q9FFU7	Q9ffu7 arabidopsis
62	7	4.3	478	11	Q9D306	Q9d306 mus musculu
63	7	4.3	478	11	Q9D2X2	Q9d2x2 mus musculu
64	7	4.3	500	17	Q978P9	Q978p9 thermoplasm
65	7	4.3	523	2	Q9KHJ9	Q9khj9 streptomyce
66	7	4.3	524	10	Q942P2	Q942p2 oryza sativ
67	7	4.3	533	16	Q8U6S7	Q8u6s7 agrobacteri
68	7	4.3	569	10	Q9FY79	Q9fy79 arabidopsis
69	7	4.3	590	16	Q9RWT8	Q9rwt8 deinococcus
70	7	4.3	604	5	Q8SYL2	Q8syl2 drosophila
71	7	4.3	629	13	Q90WR5	Q90wr5 lampetra fl
72	7	4.3	650	17	Q9HP27	Q9hp27 halobacteri
73	7	4.3	656	17	Q97AI4	Q97ai4 thermoplasm
74	7	4.3	687	16	P72758	P72758 synechocyst
75	7	4.3	696	5	Q18344	Q18344 caenorhabdi
76	7	4.3	707	2	Q9X5B5	Q9x5b5 neisseria m
77	7	4.3	714	16	Q8XV60	Q8xv60 ralstonia s
78	7	4.3	722	2	Q9X5B4	Q9x5b4 neisseria m
79	7	4.3	741	16	Q9JTK3	Q9jtk3 neisseria m
80	7	4.3	761	16	Q8X893	Q8x893 escherichia
81	7	4.3	839	16	Q97KK0	Q97kk0 clostridium
82	7	4.3	866	16	Q9L1J3	Q9l1j3 streptomyce
83	7	4.3	1001	10	Q9MAT0	Q9mat0 arabidopsis
84	7	4.3	1004	10	O48844	O48844 arabidopsis
85	7	4.3	1040	16	Q8ZNQ2	Q8znq2 salmonella
86	7	4.3	1040	16	Q8Z5F7	Q8z5f7 salmonella
87	7	4.3	1040	16	Q8X7J4	Q8x7j4 escherichia
88	7	4.3	1041	16	Q8XQM4	Q8xqm4 ralstonia s
89	7	4.3	1043	16	Q9I0V6	Q9i0v6 pseudomonas

90	7	4.3	1052	16	Q8ZCW0	Q8zcw0 yersinia pe
91	7	4.3	1061	5	Q9VXX7	Q9vxx7 drosophila
92	7	4.3	1233	5	Q9NJQ0	Q9njq0 antheraea p
93	7	4.3	1939	16	Q97DT3	Q97dt3 clostridium
94	7	4.3	2163	5	Q9NFB6	Q9nfb6 plasmodium
95	7	4.3	3263	5	Q9I7U3	Q9i7u3 drosophila
96	7	4.3	3321	4	Q8WWW7	Q8www7 homo sapien
97	7	4.3	16215	5	Q9NFS3	Q9nfs3 drosophila
98	6	3.7	28	13	Q9PTA2	Q9pta2 bungarus mu
99	6	3.7	29	11	Q62299	Q62299 mus musculu
100	6	3.7	40	4	P78341	P78341 homo sapien
101	6	3.7	40	4	P78342	P78342 homo sapien
102	6	3.7	46	2	Q93F01	Q93f01 shigella fl
103	6	3.7	46	6	Q9MZH5	Q9mzh5 callithrix
104	6	3.7	46	6	Q9MZH4	Q9mzh4 callithrix
105	6	3.7	46	6	Q9MZH3	Q9mzh3 callithrix
106	6	3.7	48	6	Q95J42	Q95j42 propithecus
107	6	3.7	54	12	Q8QLX5	Q8qlx5 influenza a
108	6	3.7	56	12	Q9PYX3	Q9pyx3 xestia c-ni
109	6	3.7	60	16	Q9PG75	Q9pg75 xylella fas
110	6	3.7	61	12	Q910Z5	Q910z5 human adeno
111	6	3.7	61	12	Q910M3	Q910m3 human adeno
112	6	3.7	63	5	Q9BJJ0	Q9bjj0 plasmodium
113	6	3.7	63	5	Q9BJI9	Q9bji9 plasmodium
114	6	3.7	64	5	Q9BJI6	Q9bji6 plasmodium
115	6	3.7	66	5	Q8SY56	Q8sy56 drosophila
116	6	3.7	66	8	Q9MHY3	Q9mhy3 pelargonium
117	6	3.7	66	8	Q9MHX3	Q9mhx3 pelargonium
118	6	3.7	66	8	Q9MHX2	Q9mhx2 pelargonium
119	6	3.7	66	8	Q9MHW8	Q9mhw8 pelargonium
120	6	3.7	71	17	Q8ZZ59	Q8zz59 pyrobaculum
121	6	3.7	73	10	Q39312	Q39312 brassica na
122	6	3.7	79	5	Q9N7T5	Q9n7t5 leishmania
123	6	3.7	82	13	Q9PTA3	Q9pta3 bungarus mu
124	6	3.7	83	12	Q8QV38	Q8qv38 avian reovi
125	6	3.7	83	12	Q8QV30	Q8qv30 avian reovi
126	6	3.7	83	12	Q8QV25	Q8qv25 avian reovi
127	6	3.7	83	12	Q8QV24	Q8qv24 avian reovi
128	6	3.7	83	13	Q90WA1	Q90wa1 pseudonaja
129	6	3.7	83	13	Q90WA0	Q90wa0 pseudonaja
130	6	3.7	83	13	Q90W99	Q90w99 pseudonaja
131	6	3.7	83	13	Q90W98	Q90w98 pseudonaja
132	6	3.7	83	13	Q90W97	Q90w97 pseudonaja
133	6	3.7	83	13	Q90W96	Q90w96 pseudonaja
134	6	3.7	86	12	Q8V9R8	Q8v9r8 african swi
135	6	3.7	88	16	Q8RB28	Q8rb28 thermoanaer
136	6	3.7	90	12	O41048	O41048 paramecium
137	6	3.7	91	12	Q9WN26	Q9wn26 hepatitis c
138	6	3.7	91	17	O27563	O27563 methanobact
139	6	3.7	93	2	O83013	O83013 streptomyce
140	6	3.7	94	16	O31537	O31537 bacillus su
141	6	3.7	96	4	Q9NTL1	Q9ntl1 homo sapien
142	6	3.7	98	13	O93292	O93292 brachydanio
143	6	3.7	100	17	Q9YG98	Q9yg98 aeropyrum p
144	6	3.7	109	13	Q9I998	Q9i998 acipenser b
145	6	3.7	111	8	Q36101	Q36101 trimorphomy
146	6	3.7	111	16	Q8ZPL2	Q8zpl2 salmonella
147	6	3.7	111	16	Q8Z6S2	Q8z6s2 salmonella
148	6	3.7	111	17	O58894	O58894 pyrococcus
149	6	3.7	111	17	Q9UZS8	Q9uzs8 pyrococcus
150	6	3.7	114	3	Q8TG09	Q8tg09 aspergillus
151	6	3.7	115	16	Q9CDX9	Q9cdx9 lactococcus
152	6	3.7	116	12	Q83599	Q83599 mumps virus
153	6	3.7	116	12	Q83601	Q83601 mumps virus
154	6	3.7	116	17	Q8TWD7	Q8twd7 methanopyru
155	6	3.7	117	5	Q9N5W2	Q9n5w2 caenorhabdi
156	6	3.7	118	16	Q9PJQ8	Q9pjq8 chlamydia m
157	6	3.7	119	16	Q97JW0	Q97jw0 clostridium
158	6	3.7	123	6	Q9GMX9	Q9gmx9 macaca fasc
159	6	3.7	124	16	Q8RAW3	Q8raw3 thermoanaer
160	6	3.7	126	2	Q9AEZ4	Q9aez4 frankia sp.
161	6	3.7	126	6	Q28878	Q28878 macaca fasc
162	6	3.7	126	6	Q28879	Q28879 macaca fasc
163	6	3.7	129	16	Q8XZ49	Q8xz49 ralstonia s
164	6	3.7	130	16	Q9PM82	Q9pm82 campylobact
165	6	3.7	131	12	Q67729	Q67729 human adeno
166	6	3.7	131	12	Q91PL2	Q91pl2 human adeno
167	6	3.7	131	16	Q97EK4	Q97ek4 clostridium
168	6	3.7	134	2	Q9X945	Q9x945 streptomyce
169	6	3.7	134	16	Q9WZH6	Q9wzh6 thermotoga
170	6	3.7	136	2	Q53355	Q53355 bacillus su
171	6	3.7	136	16	Q8XTF3	Q8xtf3 ralstonia s
172	6	3.7	137	13	Q9I999	Q9i999 acipenser b
173	6	3.7	138	12	Q8VAS2	Q8vas2 white spot
174	6	3.7	139	16	Q55975	Q55975 synechocyst
175	6	3.7	140	5	Q9BKK9	Q9bkk9 haliotis as
176	6	3.7	140	12	Q9J891	Q9j891 spodoptera
177	6	3.7	142	4	Q9H1K2	Q9h1k2 homo sapien
178	6	3.7	143	9	Q9ZXA7	Q9zxa7 bacteriopha
179	6	3.7	144	5	Q95WZ3	Q95wz3 ixodes scap
180	6	3.7	144	10	Q9SYF2	Q9syf2 arabidopsis
181	6	3.7	144	12	Q65691	Q65691 cucurbit ye
182	6	3.7	145	12	O11284	O11284 lumpy skin
183	6	3.7	145	12	Q9EEG3	Q9eeg3 bovine aden
184	6	3.7	145	16	Q9A5E3	Q9a5e3 caulobacter
185	6	3.7	147	5	Q8SU34	Q8su34 encephalito
186	6	3.7	147	16	Q927B7	Q927b7 listeria in
187	6	3.7	147	16	Q8Y3V7	Q8y3v7 listeria mo
188	6	3.7	149	6	Q9BGT5	Q9bgt5 macaca fasc
189	6	3.7	150	2	Q9L3N1	Q9l3n1 microcystis
190	6	3.7	151	17	Q8TWW2	Q8tww2 methanopyru
191	6	3.7	152	4	Q8WVK9	Q8wvk9 homo sapien
192	6	3.7	153	10	Q9SZL0	Q9szl0 arabidopsis
193	6	3.7	154	2	O88160	O88160 rhodobacter
194	6	3.7	154	5	Q8T6I1	Q8t6i1 rhipicephal
195	6	3.7	154	16	Q8U5K4	Q8u5k4 agrobacteri
196	6	3.7	155	5	O45592	O45592 caenorhabdi
197	6	3.7	155	10	Q8W3I0	Q8w3i0 oryza sativ
198	6	3.7	156	5	Q9XTV4	Q9xtv4 caenorhabdi
199	6	3.7	157	16	Q9I716	Q9i716 pseudomonas
200	6	3.7	160	8	Q8WAF9	Q8waf9 sphaerospir
201	6	3.7	160	11	Q9CRT2	Q9crt2 mus musculu
202	6	3.7	160	16	Q92WL4	Q92wl4 rhizobium m
203	6	3.7	161	16	Q9PCH9	Q9pch9 xylella fas
204	6	3.7	161	16	Q984G9	Q984g9 rhizobium l
205	6	3.7	162	2	Q9APP8	Q9app8 uncultured
206	6	3.7	162	12	Q66529	Q66529 equine arte
207	6	3.7	162	12	Q66515	Q66515 equine arte
208	6	3.7	162	12	Q9YNV0	Q9ynv0 equine arte
209	6	3.7	162	12	Q9YNT8	Q9ynt8 equine arte
210	6	3.7	162	12	Q66516	Q66516 equine arte
211	6	3.7	162	12	Q9WA50	Q9wa50 equine arte
212	6	3.7	162	12	Q9WA04	Q9wa04 equine arte
213	6	3.7	163	17	Q9YD34	Q9yd34 aeropyrum p
214	6	3.7	164	16	Q9PNA4	Q9pna4 campylobact
215	6	3.7	165	1	Q49172	Q49172 methanother
216	6	3.7	165	5	Q9NFW1	Q9nfw1 hydra atten
217	6	3.7	165	17	O28330	O28330 archaeoglob
218	6	3.7	167	2	Q9X2X2	Q9x2x2 bacillus an
219	6	3.7	167	13	Q9DET4	Q9det4 myoxocephal
220	6	3.7	167	16	Q98BQ5	Q98bq5 rhizobium l
221	6	3.7	169	16	Q8RBD2	Q8rbd2 thermoanaer
222	6	3.7	172	16	Q8ZQG1	Q8zqg1 salmonella
223	6	3.7	172	16	Q8XUD7	Q8xud7 ralstonia s
224	6	3.7	174	5	O76968	O76968 podocoryne
225	6	3.7	174	10	Q8SAW0	Q8saw0 oryza sativ
226	6	3.7	178	5	Q8SVH2	Q8svh2 encephalito
227	6	3.7	178	16	Q9ZJK6	Q9zjk6 helicobacte
228	6	3.7	179	16	Q8YX74	Q8yx74 anabaena sp
229	6	3.7	184	11	Q9JK63	Q9jk63 mus musculu
230	6	3.7	185	10	Q948R3	Q948r3 oryza sativ
231	6	3.7	185	16	Q8YW51	Q8yw51 anabaena sp
232	6	3.7	186	12	Q91NR1	Q91nr1 helicoverpa
233	6	3.7	187	2	Q9KK58	Q9kk58 streptococc
234	6	3.7	187	2	Q53571	Q53571 synechococc
235	6	3.7	188	16	Q9L1G0	Q9l1g0 streptomyce

236	6	3.7	191	6	O97635	O97635 pithecia pi
237	6	3.7	191	6	O97636	O97636 pithecia pi
238	6	3.7	191	6	O97637	O97637 alouatta se
239	6	3.7	191	6	O97638	O97638 alouatta se
240	6	3.7	191	6	O97639	O97639 pithecia pi
241	6	3.7	191	6	Q9TV73	Q9tv73 cebus oliva
242	6	3.7	191	16	Q92P73	Q92p73 rhizobium m
243	6	3.7	193	2	Q9EV55	Q9ev55 rhizobium m
244	6	3.7	194	2	Q9RQQ6	Q9rqq6 bacillus am
245	6	3.7	194	10	Q9LFV1	Q9lfv1 arabidopsis
246	6	3.7	194	16	Q8REJ7	Q8rej7 fusobacteri
247	6	3.7	195	5	O45149	O45149 caenorhabdi
248	6	3.7	195	16	Q92KD7	Q92kd7 rhizobium m
249	6	3.7	195	16	Q9RKD9	Q9rkd9 streptomyce
250	6	3.7	196	8	O79232	O79232 nucella ema
251	6	3.7	196	8	Q9T8Z0	Q9t8z0 nucella ost
252	6	3.7	196	8	Q9T8Y4	Q9t8y4 nucella ema
253	6	3.7	196	8	Q9T8Y3	Q9t8y3 nucella ema
254	6	3.7	196	8	Q9T8Y2	Q9t8y2 nucella ost
255	6	3.7	196	8	Q9T8Y1	Q9t8y1 nucella ema
256	6	3.7	196	8	Q9T4I5	Q9t4i5 nucella ema
257	6	3.7	196	8	Q9T4F0	Q9t4f0 nucella ema
258	6	3.7	198	5	Q9GR68	Q9gr68 caenorhabdi
259	6	3.7	199	6	Q9TSD6	Q9tsd6 callithrix
260	6	3.7	199	10	Q8SA98	Q8sa98 zea mays (m
261	6	3.7	201	16	O34943	O34943 bacillus su
262	6	3.7	202	5	Q22473	Q22473 caenorhabdi
263	6	3.7	202	16	Q8ZH93	Q8zh93 yersinia pe
264	6	3.7	202	17	Q8ZW61	Q8zw61 pyrobaculum
265	6	3.7	205	10	Q9ZPM1	Q9zpm1 eucalyptus
266	6	3.7	205	16	Q9PNH9	Q9pnh9 campylobact
267	6	3.7	205	16	Q8Y1G3	Q8y1g3 ralstonia s
268	6	3.7	206	2	O52152	O52152 escherichia
269	6	3.7	206	16	Q9KZ92	Q9kz92 streptomyce
270	6	3.7	208	8	Q8WEQ0	Q8weq0 lapemis har
271	6	3.7	208	16	Q92QA2	Q92qa2 rhizobium m
272	6	3.7	209	12	Q9YS13	Q9ys13 norwalk vir
273	6	3.7	210	16	Q8XI88	Q8xi88 clostridium
274	6	3.7	213	16	Q8ZPP1	Q8zpp1 salmonella
275	6	3.7	213	16	Q8Z6N8	Q8z6n8 salmonella
276	6	3.7	213	16	Q8YGR1	Q8ygr1 brucella me
277	6	3.7	216	12	Q84523	Q84523 paramecium
278	6	3.7	216	16	Q8Z566	Q8z566 salmonella
279	6	3.7	216	16	Q8YPU5	Q8ypu5 anabaena sp
280	6	3.7	217	5	Q963B6	Q963b6 spodoptera
281	6	3.7	217	8	O47530	O47530 timema rite
282	6	3.7	217	16	O67896	O67896 aquifex aeo
283	6	3.7	219	2	P74887	P74887 salmonella
284	6	3.7	219	10	O82268	O82268 arabidopsis
285	6	3.7	219	16	Q8XEP7	Q8xep7 salmonella
286	6	3.7	221	16	Q9CEF4	Q9cef4 lactococcus
287	6	3.7	221	16	Q985A6	Q985a6 rhizobium l
288	6	3.7	222	2	Q93JP6	Q93jp6 ruminococcu
289	6	3.7	222	12	Q9YMQ2	Q9ymq2 lymantria d
290	6	3.7	223	10	Q9FSN2	Q9fsn2 oryza sativ
291	6	3.7	224	9	O64370	O64370 lactobacill
292	6	3.7	224	12	Q9PYX2	Q9pyx2 xestia c-ni
293	6	3.7	224	16	Q99ZG8	Q99zg8 streptococc
294	6	3.7	225	12	P87635	P87635 cucurbit ye
295	6	3.7	225	13	Q9PTT2	Q9ptt2 lepidobatra
296	6	3.7	225	17	Q8ZYC7	Q8zyc7 pyrobaculum
297	6	3.7	226	5	Q8SVS2	Q8svs2 encephalito
298	6	3.7	226	16	Q9CFK2	Q9cfk2 lactococcus
299	6	3.7	227	4	Q9NSM3	Q9nsm3 homo sapien
300	6	3.7	227	16	Q8ZA57	Q8za57 yersinia pe
301	6	3.7	228	16	Q99YN1	Q99yn1 streptococc
302	6	3.7	230	5	Q9VS96	Q9vs96 drosophila
303	6	3.7	230	12	O90313	O90313 chlorella v
304	6	3.7	230	16	Q92KT2	Q92kt2 rhizobium m
305	6	3.7	232	16	Q9RXW0	Q9rxw0 deinococcus
306	6	3.7	234	5	Q9VTP4	Q9vtp4 drosophila
307	6	3.7	235	10	Q43273	Q43273 zea mays (m
308	6	3.7	235	17	Q9V1D8	Q9v1d8 pyrococcus
309	6	3.7	236	5	Q95RI6	Q95ri6 drosophila
310	6	3.7	236	5	Q95SS8	Q95ss8 drosophila
311	6	3.7	237	6	Q9BDH0	Q9bdh0 canis famil
312	6	3.7	238	12	Q91TU5	Q91tu5 tupaia herp
313	6	3.7	239	10	Q94DZ8	Q94dz8 oryza sativ
314	6	3.7	239	16	O53242	O53242 mycobacteri
315	6	3.7	240	10	Q9SRN1	Q9srn1 arabidopsis
316	6	3.7	240	10	Q9LKQ3	Q9lkq3 arabidopsis
317	6	3.7	240	17	Q8U1X4	Q8u1x4 pyrococcus
318	6	3.7	242	2	O67989	O67989 rhodococcus
319	6	3.7	242	8	Q94UZ5	Q94uz5 leishmania
320	6	3.7	242	17	Q9HJE3	Q9hje3 thermoplasm
321	6	3.7	242	17	O27260	O27260 methanobact
322	6	3.7	243	16	Q97GL9	Q97gl9 clostridium
323	6	3.7	243	16	O53733	O53733 mycobacteri
324	6	3.7	243	17	Q8ZV99	Q8zv99 pyrobaculum
325	6	3.7	244	16	Q9KYS4	Q9kys4 streptomyce
326	6	3.7	247	3	Q9UT40	Q9ut40 schizosacch
327	6	3.7	248	8	Q9MI65	Q9mi65 crithidia f
328	6	3.7	248	10	Q9FLU4	Q9flu4 arabidopsis
329	6	3.7	248	11	Q9DD10	Q9dd10 mus musculu
330	6	3.7	248	16	O85622	O85622 escherichia
331	6	3.7	250	5	Q9TYP7	Q9typ7 caenorhabdi
332	6	3.7	250	8	Q9TC90	Q9tc90 nephroselmi
333	6	3.7	250	8	Q9G8P2	Q9g8p2 naegleria g
334	6	3.7	250	17	Q9HJ20	Q9hj20 thermoplasm
335	6	3.7	251	16	Q985W4	Q985w4 rhizobium l
336	6	3.7	252	15	P89687	P89687 feline immu
337	6	3.7	252	16	Q9HVC9	Q9hvc9 pseudomonas
338	6	3.7	252	16	Q9A5Z1	Q9a5z1 caulobacter
339	6	3.7	254	16	P74487	P74487 synechocyst
340	6	3.7	255	16	Q8ZRV3	Q8zrv3 salmonella
341	6	3.7	255	16	Q8Z9I7	Q8z917 salmonella
342	6	3.7	255	16	Q8YD43	Q8yd43 brucella me
343	6	3.7	256	2	P97060	P97060 streptomyce
344	6	3.7	256	16	O67182	O67182 aquifex aeo
345	6	3.7	257	16	Q92DK8	Q92dk8 listeria in
346	6	3.7	258	10	Q9XIC2	Q9xic2 arabidopsis
347	6	3.7	258	16	Q9HY97	Q9hy97 pseudomonas
348	6	3.7	259	16	Q55831	Q55831 synechocyst
349	6	3.7	260	3	Q9UUB8	Q9uub8 schizosacch
350	6	3.7	261	2	Q55061	Q55061 synechocyst
351	6	3.7	261	11	Q9DAF0	Q9daf0 mus musculu
352	6	3.7	261	16	P73414	P73414 synechocyst
353	6	3.7	261	16	P73683	P73683 synechocyst
354	6	3.7	261	16	P73880	P73880 synechocyst
355	6	3.7	261	16	Q8XBA1	Q8xba1 escherichia
356	6	3.7	261	16	Q57495	Q57495 synechocyst
357	6	3.7	262	2	Q9RB26	Q9rb26 pectobacter
358	6	3.7	262	2	Q9XB51	Q9xb51 erwinia car
359	6	3.7	262	5	Q9TZA1	Q9tza1 caenorhabdi
360	6	3.7	262	16	Q9CIQ8	Q9ciq8 lactococcus
361	6	3.7	263	16	P73009	P73009 synechocyst
362	6	3.7	264	16	Q9KTA4	Q9kta4 vibrio chol
363	6	3.7	265	13	Q90WV8	Q90wv8 xenopus lae
364	6	3.7	266	16	Q9PBA0	Q9pba0 xylella fas
365	6	3.7	267	16	Q92S53	Q92s53 rhizobium m
366	6	3.7	268	16	Q8RGZ5	Q8rgz5 fusobacteri
367	6	3.7	270	10	O65489	O65489 arabidopsis
368	6	3.7	270	16	Q9CMH3	Q9cmh3 pasteurella
369	6	3.7	273	5	Q9VZI5	Q9vzi5 drosophila
370	6	3.7	273	10	Q9LK01	Q9lk01 arabidopsis
371	6	3.7	273	16	Q8XXY5	Q8xxy5 ralstonia s
372	6	3.7	275	5	Q20234	Q20234 caenorhabdi
373	6	3.7	275	10	O81801	O81801 arabidopsis
374	6	3.7	275	16	Q99UZ5	Q99uz5 staphylococ
375	6	3.7	276	5	Q18061	Q18061 caenorhabdi
376	6	3.7	276	11	Q9WTY0	Q9wty0 rattus norv
377	6	3.7	276	11	Q9D0Z3	Q9d0z3 mus musculu
378	6	3.7	276	16	Q9CP57	Q9cp57 pasteurella
379	6	3.7	278	10	Q9SUL3	Q9sul3 arabidopsis
380	6	3.7	279	2	Q52937	Q52937 rhizobium m
381	6	3.7	279	8	Q94XY6	Q94xy6 fenneropena

382	6	3.7	279	8	Q94P15	Q94p15 fenneropena
383	6	3.7	280	6	O97624	O97624 canis famil
384	6	3.7	280	16	Q98GY3	Q98gy3 rhizobium l
385	6	3.7	283	2	O31397	O31397 bacillus st
386	6	3.7	283	2	Q45937	Q45937 coxiella bu
387	6	3.7	283	11	Q8VDW5	Q8vdw5 mus musculu
388	6	3.7	284	4	Q99781	Q99781 homo sapien
389	6	3.7	286	16	Q8ZQD9	Q8zqd9 salmonella
390	6	3.7	286	16	Q8Z818	Q8z818 salmonella
391	6	3.7	287	4	Q9BR04	Q9br04 homo sapien
392	6	3.7	289	5	Q9NEP2	Q9nep2 caenorhabdi
393	6	3.7	289	17	Q96Z24	Q96z24 sulfolobus
394	6	3.7	290	4	Q9UQS0	Q9uqs0 homo sapien
395	6	3.7	291	16	Q92U55	Q92u55 rhizobium m
396	6	3.7	291	16	Q10828	Q10828 mycobacteri
397	6	3.7	291	17	Q9HIY1	Q9hiy1 thermoplasm
398	6	3.7	292	16	Q97KR9	Q97kr9 clostridium
399	6	3.7	293	5	Q9U1R9	Q9u1r9 caenorhabdi
400	6	3.7	293	10	Q9SLV9	Q9slv9 cucumis sat
401	6	3.7	293	13	Q9DEX9	Q9dex9 cyprinus ca
402	6	3.7	293	16	Q98LU1	Q98lu1 rhizobium l
403	6	3.7	293	16	Q9KY30	Q9ky30 streptomyce
404	6	3.7	294	10	Q39653	Q39653 cucumis sat
405	6	3.7	294	17	Q8TX11	Q8tx11 methanopyru
406	6	3.7	295	16	Q92M02	Q92m02 rhizobium m
407	6	3.7	296	2	Q45339	Q45339 bordetella
408	6	3.7	296	16	Q92ZH0	Q92zh0 rhizobium m
409	6	3.7	296	17	Q8TU61	Q8tu61 methanosarc
410	6	3.7	297	10	Q9SA66	Q9sa66 arabidopsis
411	6	3.7	297	17	Q9HQ93	Q9hq93 halobacteri
412	6	3.7	298	2	Q49806	Q49806 mycobacteri
413	6	3.7	298	17	Q97UA0	Q97ua0 sulfolobus
414	6	3.7	299	5	O17255	O17255 caenorhabdi
415	6	3.7	300	4	Q96EF3	Q96ef3 homo sapien
416	6	3.7	300	16	Q9L231	Q9l231 streptomyce
417	6	3.7	301	2	Q93DN7	Q93dn7 uncultured
418	6	3.7	301	5	Q9VUN2	Q9vun2 drosophila
419	6	3.7	301	16	O33234	O33234 mycobacteri
420	6	3.7	301	16	Q8X4Y0	Q8x4y0 escherichia
421	6	3.7	302	1	Q9C4R6	Q9c4r6 methanosarc
422	6	3.7	302	2	Q93L28	Q93l28 burkholderi
423	6	3.7	302	10	Q93Y00	Q93y00 arabidopsis
424	6	3.7	303	6	Q95KC6	Q95kc6 macaca fasc
425	6	3.7	303	16	Q97GQ4	Q97gq4 clostridium
426	6	3.7	303	16	Q8YHF8	Q8yhf8 brucella me
427	6	3.7	304	2	Q9XDS0	Q9xds0 lactobacill
428	6	3.7	305	2	Q9F132	Q9f132 pseudomonas
429	6	3.7	305	13	O73743	O73743 xenopus lae
430	6	3.7	305	17	O30284	O30284 archaeoglob
431	6	3.7	306	16	Q98L61	Q98l61 rhizobium l
432	6	3.7	306	16	Q8RAR8	Q8rar8 thermoanaer
433	6	3.7	306	17	Q9YCK4	Q9yck4 aeropyrum p
434	6	3.7	307	10	O65737	O65737 cicer ariet
435	6	3.7	307	16	Q9F352	Q9f352 streptomyce
436	6	3.7	308	10	Q9ZU73	Q9zu73 arabidopsis
437	6	3.7	308	16	Q92LR4	Q92lr4 rhizobium m
438	6	3.7	309	16	Q9X019	Q9x019 thermotoga
439	6	3.7	310	16	Q9I1D9	Q9i1d9 pseudomonas
440	6	3.7	310	16	Q984X2	Q984x2 rhizobium l
441	6	3.7	312	16	Q8Y312	Q8y312 ralstonia s
442	6	3.7	313	16	Q8ZB10	Q8zb10 yersinia pe
443	6	3.7	314	4	Q9Y3B0	Q9y3b0 homo sapien
444	6	3.7	314	4	Q96GK7	Q96gk7 homo sapien
445	6	3.7	314	10	Q9LVZ3	Q9lvz3 arabidopsis
446	6	3.7	314	16	Q8ZPL1	Q8zpl1 salmonella
447	6	3.7	314	16	Q8Z6X4	Q8z6x4 salmonella
448	6	3.7	314	16	Q8X773	Q8x773 escherichia
449	6	3.7	314	16	Q92MS1	Q92ms1 rhizobium m
450	6	3.7	315	4	Q96JZ6	Q96jz6 homo sapien
451	6	3.7	315	4	Q96SU7	Q96su7 homo sapien
452	6	3.7	315	6	Q9BE05	Q9be05 macaca fasc
453	6	3.7	315	6	Q95J75	Q95j75 macaca fasc
454	6	3.7	315	16	Q98JF7	Q98jf7 rhizobium l
455	6	3.7	315	17	Q9HK55	Q9hk55 thermoplasm
456	6	3.7	317	2	O05214	O05214 bacillus st
457	6	3.7	317	16	Q8ZE77	Q8ze77 yersinia pe
458	6	3.7	318	5	O76194	O76194 drosophila
459	6	3.7	318	16	Q9KMZ4	Q9kmz4 vibrio chol
460	6	3.7	318	16	Q8U9A7	Q8u9a7 agrobacteri
461	6	3.7	318	17	Q9HRU5	Q9hru5 halobacteri
462	6	3.7	319	2	Q8VLZ1	Q8vlz1 treponema d
463	6	3.7	319	16	Q8XYD4	Q8xyd4 ralstonia s
464	6	3.7	319	16	Q8UDA2	Q8uda2 agrobacteri
465	6	3.7	322	5	Q8SXD9	Q8sxd9 drosophila
466	6	3.7	322	13	Q90XB0	Q90xb0 petromyzon
467	6	3.7	323	2	O68908	O68908 streptomyce
468	6	3.7	324	9	Q38582	Q38582 bacteriopha
469	6	3.7	324	9	Q8SCY9	Q8scy9 pseudomonas
470	6	3.7	325	16	Q9A316	Q9a316 caulobacter
471	6	3.7	325	16	Q8XVG2	Q8xvg2 ralstonia s
472	6	3.7	326	3	Q9P7Q2	Q9p7q2 schizosacch
473	6	3.7	326	12	Q9J862	Q9j862 spodoptera
474	6	3.7	326	12	Q8QV17	Q8qv17 avian reovi
475	6	3.7	326	12	Q8QV13	Q8qv13 avian reovi
476	6	3.7	326	12	Q8QV12	Q8qv12 avian reovi
477	6	3.7	326	16	Q8XAU2	Q8xau2 escherichia
478	6	3.7	327	5	O97190	O97190 geodia cydo
479	6	3.7	333	16	Q8YVW1	Q8yvw1 anabaena sp
480	6	3.7	334	2	Q9ZGQ3	Q9zgq3 burkholderi
481	6	3.7	334	16	Q9AK31	Q9ak31 streptomyce
482	6	3.7	334	17	Q8U0R7	Q8u0r7 pyrococcus
483	6	3.7	335	16	Q9KNQ4	Q9knq4 vibrio chol
484	6	3.7	336	5	Q9NMI6	Q9nmi6 leishmania
485	6	3.7	336	16	Q9JYZ6	Q9jyz6 neisseria m
486	6	3.7	336	16	Q8U8M7	Q8u8m7 agrobacteri
487	6	3.7	338	5	O17925	O17925 caenorhabdi
488	6	3.7	338	17	O29023	O29023 archaeoglob
489	6	3.7	339	3	O14421	O14421 cochliobolu
490	6	3.7	340	3	Q9UV52	Q9uv52 ustilago ho
491	6	3.7	340	16	Q55632	Q55632 synechocyst
492	6	3.7	340	17	Q9HIL1	Q9hil1 thermoplasm
493	6	3.7	341	17	Q8THW7	Q8thw7 methanosarc
494	6	3.7	342	2	P96096	P96096 thiobacillu
495	6	3.7	342	7	Q95556	Q95556 peromyscus
496	6	3.7	344	5	Q95Q38	Q95q38 caenorhabdi
497	6	3.7	344	17	Q97Z28	Q97z28 sulfolobus
498	6	3.7	345	5	Q9VXR2	Q9vxr2 drosophila
499	6	3.7	345	10	Q9M9Y4	Q9m9y4 arabidopsis
500	6	3.7	345	16	Q8ZAN4	Q8zan4 yersinia pe
501	6	3.7	345	17	Q8U0F1	Q8u0f1 pyrococcus
502	6	3.7	346	5	Q8T0F8	Q8t0f8 drosophila
503	6	3.7	347	11	Q9DBV7	Q9dbv7 mus musculu
504	6	3.7	347	16	Q92MV7	Q92mv7 rhizobium m
505	6	3.7	347	16	Q8YIK9	Q8yik9 brucella me
506	6	3.7	349	5	Q95WF1	Q95wf1 drosophila
507	6	3.7	349	17	O59613	O59613 pyrococcus
508	6	3.7	350	5	Q9V5E3	Q9v5e3 drosophila
509	6	3.7	350	10	Q9LDR8	Q9ldr8 arabidopsis
510	6	3.7	350	10	Q8W4E0	Q8w4e0 arabidopsis
511	6	3.7	350	16	Q9X1B5	Q9x1b5 thermotoga
512	6	3.7	351	16	Q9A2X2	Q9a2x2 caulobacter
513	6	3.7	351	17	Q980F5	Q980f5 sulfolobus
514	6	3.7	352	2	O68498	O68498 pseudoalter
515	6	3.7	352	10	O65409	O65409 arabidopsis
516	6	3.7	353	16	Q8YZN1	Q8yzn1 anabaena sp
517	6	3.7	353	16	Q9F3B4	Q9f3b4 streptomyce
518	6	3.7	355	17	O27424	O27424 methanobact
519	6	3.7	356	5	Q95WF3	Q95wf3 drosophila
520	6	3.7	356	5	Q8SY01	Q8sy01 drosophila
521	6	3.7	356	10	Q9LQQ3	Q9lqq3 arabidopsis
522	6	3.7	356	16	Q8U934	Q8u934 agrobacteri
523	6	3.7	357	12	Q91TU3	Q91tu3 tupaia herp
524	6	3.7	358	3	Q9P5N1	Q9p5n1 schizosacch
525	6	3.7	358	4	Q8TEI8	Q8tei8 homo sapien
526	6	3.7	358	16	Q8XPB2	Q8xpb2 clostridium
527	6	3.7	360	5	Q9U6V0	Q9u6v0 ciona intes

528	6	3.7	360	16	Q92LH8	Q92lh8 rhizobium m
529	6	3.7	361	5	Q94626	Q94626 naegleria f
530	6	3.7	361	12	Q99GV5	Q99gv5 helicoverpa
531	6	3.7	361	16	Q8XVN3	Q8xvn3 ralstonia s
532	6	3.7	364	6	O77567	O77567 saimiri sci
533	6	3.7	364	6	O77568	O77568 saimiri sci
534	6	3.7	364	6	O77569	O77569 saimiri sci
535	6	3.7	364	6	O77572	O77572 saguinus my
536	6	3.7	364	6	O77573	O77573 saguinus my
537	6	3.7	364	6	O77574	O77574 saguinus my
538	6	3.7	364	6	O77575	O77575 cebus oliva
539	6	3.7	364	6	O77576	O77576 cebus oliva
540	6	3.7	364	6	Q9TSH3	Q9tsh3 cebus oliva
541	6	3.7	364	6	O62799	O62799 tursiops tr
542	6	3.7	364	6	O97901	O97901 phoca vitul
543	6	3.7	364	6	Q9XSF1	Q9xsf1 odocoileus
544	6	3.7	364	6	Q9TU70	Q9tu70 equus cabal
545	6	3.7	364	6	Q9XSX4	Q9xsx4 macaca fasc
546	6	3.7	364	6	Q9XSX3	Q9xsx3 macaca fasc
547	6	3.7	364	6	Q9XSX2	Q9xsx2 macaca fasc
548	6	3.7	364	6	Q9BGI7	Q9bgi7 bos taurus
549	6	3.7	364	6	Q9TSH4	Q9tsh4 callithrix
550	6	3.7	364	6	O77570	O77570 callithrix
551	6	3.7	364	6	O77571	O77571 callithrix
552	6	3.7	364	6	O62860	O62860 otolemur ga
553	6	3.7	364	16	Q92RX4	Q92rx4 rhizobium m
554	6	3.7	365	5	Q8T750	Q8t750 branchiosto
555	6	3.7	366	16	Q8UC47	Q8uc47 agrobacteri
556	6	3.7	367	10	Q9SV64	Q9sv64 arabidopsis
557	6	3.7	367	16	Q9WYR4	Q9wyr4 thermotoga
558	6	3.7	370	3	Q04083	Q04083 saccharomyc
559	6	3.7	371	12	Q91BJ8	Q91bj8 spodoptera
560	6	3.7	373	2	Q9ZGC0	Q9zgc0 streptomyce
561	6	3.7	375	16	O06179	O06179 mycobacteri
562	6	3.7	375	16	Q9ZLL1	Q9zll1 helicobacte
563	6	3.7	375	16	O25341	O25341 helicobacte
564	6	3.7	376	10	Q9FPH0	Q9fph0 arabidopsis
565	6	3.7	376	16	O06315	O06315 mycobacteri
566	6	3.7	377	5	Q9VB68	Q9vb68 drosophila
567	6	3.7	378	4	Q96H15	Q96h15 homo sapien
568	6	3.7	379	4	Q96P88	Q96p88 homo sapien
569	6	3.7	379	5	Q95WF2	Q95wf2 drosophila
570	6	3.7	379	6	Q95MH6	Q95mh6 cercopithec
571	6	3.7	379	6	Q95JG1	Q95jg1 macaca mula
572	6	3.7	379	10	O65542	O65542 arabidopsis
573	6	3.7	380	6	Q95MG6	Q95mg6 callithrix
574	6	3.7	380	16	Q8XXD8	Q8xxd8 ralstonia s
575	6	3.7	382	8	Q98S66	Q98s66 guillardia
576	6	3.7	383	13	Q90YY7	Q90yy7 brachydanio
577	6	3.7	383	17	O26855	O26855 methanobact
578	6	3.7	384	10	Q8RYF8	Q8ryf8 oryza sativ
579	6	3.7	386	16	Q9CIY3	Q9ciy3 lactococcus
580	6	3.7	386	17	Q96Y15	Q96y15 sulfolobus
581	6	3.7	388	10	Q39902	Q39902 glycine max
582	6	3.7	388	16	Q97F25	Q97f25 clostridium
583	6	3.7	388	17	Q8ZZ97	Q8zz97 pyrobaculum
584	6	3.7	389	4	Q96HU3	Q96hu3 homo sapien
585	6	3.7	389	16	Q8RHE9	Q8rhe9 fusobacteri
586	6	3.7	390	2	Q9R6K4	Q9r6k4 agrobacteri
587	6	3.7	390	6	Q8WNV9	Q8wnv9 sus scrofa
588	6	3.7	390	13	Q9W6U5	Q9w6u5 fugu rubrip
589	6	3.7	390	16	Q9K432	Q9k432 streptomyce
590	6	3.7	391	11	Q61177	Q61177 mus musculu
591	6	3.7	391	11	Q8R0X4	Q8r0x4 mus musculu
592	6	3.7	392	2	Q8VN08	Q8vn08 lactobacill
593	6	3.7	392	17	O28978	O28978 archaeoglob
594	6	3.7	393	8	Q9TAR6	Q9tar6 ctonoxylon
595	6	3.7	393	16	Q987L3	Q987l3 rhizobium l
596	6	3.7	394	5	O62147	O62147 caenorhabdi
597	6	3.7	394	10	Q9SYK6	Q9syk6 arabidopsis
598	6	3.7	394	10	Q9SV71	Q9sv71 arabidopsis
599	6	3.7	394	12	Q91GF9	Q91gf9 epiphyas po
600	6	3.7	394	16	Q8XL37	Q8xl37 clostridium
601	6	3.7	395	16	Q97FE2	Q97fe2 clostridium
602	6	3.7	396	2	Q937U1	Q937u1 brucella su
603	6	3.7	396	16	Q8YD15	Q8yd15 brucella me
604	6	3.7	396	16	Q8XVG0	Q8xvg0 ralstonia s
605	6	3.7	397	16	Q9RU94	Q9ru94 deinococcus
606	6	3.7	398	5	Q24466	Q24466 drosophila
607	6	3.7	398	5	Q9VQQ0	Q9vqq0 drosophila
608	6	3.7	404	16	Q9A1E0	Q9ale0 streptococc
609	6	3.7	404	16	Q8XYW7	Q8xyw7 ralstonia s
610	6	3.7	405	16	Q9JV53	Q9jv53 neisseria m
611	6	3.7	405	16	Q97H14	Q97h14 clostridium
612	6	3.7	406	4	Q9Y2U9	Q9y2u9 homo sapien
613	6	3.7	406	5	O17661	O17661 caenorhabdi
614	6	3.7	406	11	Q9D784	Q9d784 mus musculu
615	6	3.7	406	11	Q99JY2	Q99jy2 mus musculu
616	6	3.7	407	16	Q9K043	Q9k043 neisseria m
617	6	3.7	408	16	Q9RZI1	Q9rzi1 deinococcus
618	6	3.7	410	5	Q9VQ13	Q9vq13 drosophila
619	6	3.7	411	2	Q9F2I5	Q9f2i5 streptococc
620	6	3.7	411	16	Q98J65	Q98j65 rhizobium l
621	6	3.7	411	16	O54109	O54109 streptomyce
622	6	3.7	412	16	Q9PJA6	Q9pja6 campylobact
623	6	3.7	412	16	Q92HK7	Q92hk7 rickettsia
624	6	3.7	414	2	Q53038	Q53038 rhodococcus
625	6	3.7	414	5	Q21405	Q21405 caenorhabdi
626	6	3.7	414	12	Q81269	Q81269 hepatitis c
627	6	3.7	414	12	Q81270	Q81270 hepatitis c
628	6	3.7	415	5	O17406	O17406 caenorhabdi
629	6	3.7	415	16	Q9WZM0	Q9wzm0 thermotoga
630	6	3.7	416	13	Q8QHK4	Q8qhk4 lampetra ja
631	6	3.7	417	16	Q9WYU7	Q9wyu7 thermotoga
632	6	3.7	418	5	Q8T9U7	Q8t9u7 aedes aegyp
633	6	3.7	418	10	Q9FYL1	Q9fyl1 arabidopsis
634	6	3.7	418	11	O70426	O70426 rattus norv
635	6	3.7	418	16	P73784	P73784 synechocyst
636	6	3.7	419	10	Q8S1V0	Q8s1v0 oryza sativ
637	6	3.7	420	2	O68332	O68332 pseudomonas
638	6	3.7	420	9	Q8SDA7	Q8sda7 pseudomonas
639	6	3.7	420	10	Q9LIS1	Q9lis1 arabidopsis
640	6	3.7	420	10	Q94AY3	Q94ay3 arabidopsis
641	6	3.7	421	2	Q9ZNF6	Q9znf6 bacillus li
642	6	3.7	422	16	Q8UAF5	Q8uaf5 agrobacteri
643	6	3.7	423	4	Q8TE80	Q8te80 homo sapien
644	6	3.7	423	9	Q9AYZ7	Q9ayz7 bacteriopha
645	6	3.7	423	16	Q8ZIQ5	Q8ziq5 yersinia pe
646	6	3.7	424	9	Q9ZXJ8	Q9zxj8 bacteriopha
647	6	3.7	425	4	Q9BUM6	Q9bum6 homo sapien
648	6	3.7	425	16	Q99RC3	Q99rc3 staphylococ
649	6	3.7	426	16	Q9WY78	Q9wy78 thermotoga
650	6	3.7	427	16	Q8XTJ2	Q8xtj2 ralstonia s
651	6	3.7	427	17	Q8TNR5	Q8tnr5 methanosarc
652	6	3.7	428	10	Q9SVM9	Q9svm9 arabidopsis
653	6	3.7	428	13	Q9PWC8	Q9pwc8 brachydanio
654	6	3.7	428	16	Q9ABC9	Q9abc9 caulobacter
655	6	3.7	428	16	Q8ZDS9	Q8zds9 yersinia pe
656	6	3.7	428	17	Q59670	Q59670 pyrococcus
657	6	3.7	430	13	Q98SS3	Q98ss3 brachydanio
658	6	3.7	430	13	Q98SS0	Q98ss0 brachydanio
659	6	3.7	431	13	Q98SS2	Q98ss2 brachydanio
660	6	3.7	432	13	Q98SS1	Q98ss1 brachydanio
661	6	3.7	433	10	Q05929	Q05929 daucus caro
662	6	3.7	433	11	Q9QWF2	Q9qwf2 rattus sp.
663	6	3.7	434	2	O07364	O07364 chlamydia t
664	6	3.7	434	16	Q9PKT9	Q9pkt9 chlamydia m
665	6	3.7	434	16	O84099	O84099 chlamydia t
666	6	3.7	434	16	Q8XRD2	Q8xrd2 ralstonia s
667	6	3.7	435	16	Q9CH92	Q9ch92 lactococcus
668	6	3.7	437	5	Q8T0I5	Q8t0i5 drosophila
669	6	3.7	437	10	Q9M1D6	Q9m1d6 arabidopsis
670	6	3.7	437	16	Q9KYL8	Q9kyl8 streptomyce
671	6	3.7	438	16	Q8ZLI6	Q8zli6 salmonella
672	6	3.7	438	16	Q8ZA64	Q8za64 yersinia pe
673	6	3.7	438	16	Q8Z224	Q8z224 salmonella

674	6	3.7	438	16	Q8X712	Q8x712 escherichia
675	6	3.7	438	16	Q8VIX7	Q8vix7 mycobacteri
676	6	3.7	440	3	Q8TG21	Q8tg21 saccharomyc
677	6	3.7	443	16	Q9L051	Q9l051 streptomyce
678	6	3.7	445	10	Q8RYK3	Q8ryk3 oryza sativ
679	6	3.7	445	17	Q8U123	Q8u123 pyrococcus
680	6	3.7	446	4	O60650	O60650 homo sapien
681	6	3.7	446	16	Q8X3N4	Q8x3n4 escherichia
682	6	3.7	448	16	Q8YXW8	Q8yxw8 anabaena sp
683	6	3.7	449	16	Q9A3Q1	Q9a3q1 caulobacter
684	6	3.7	450	10	Q9LHE6	Q9lhe6 arabidopsis
685	6	3.7	450	16	O07752	O07752 mycobacteri
686	6	3.7	451	16	Q99ZW8	Q99zw8 streptococc
687	6	3.7	452	16	O07424	O07424 mycobacteri
688	6	3.7	452	16	O06155	O06155 mycobacteri
689	6	3.7	453	3	Q96VT0	Q96vt0 agrocybe pe
690	6	3.7	453	5	Q8T3E6	Q8t3e6 caenorhabdi
691	6	3.7	453	5	Q8SUM4	Q8sum4 encephalito
692	6	3.7	453	16	Q8U794	Q8u794 agrobacteri
693	6	3.7	454	16	Q8YKM0	Q8ykm0 anabaena sp
694	6	3.7	455	5	Q19553	Q19553 caenorhabdi
695	6	3.7	457	16	Q9HVD8	Q9hvd8 pseudomonas
696	6	3.7	458	10	Q9AYL9	Q9ayl9 oryza sativ
697	6	3.7	459	5	Q9VSS3	Q9vss3 drosophila
698	6	3.7	459	11	Q9D852	Q9d852 mus musculu
699	6	3.7	461	5	O96407	O96407 litomosoide
700	6	3.7	465	11	Q8VCM0	Q8vcm0 mus musculu
701	6	3.7	466	4	Q9NXP0	Q9nxp0 homo sapien
702	6	3.7	467	10	Q9FZK1	Q9fzk1 arabidopsis
703	6	3.7	467	10	Q9SJV9	Q9sjv9 arabidopsis
704	6	3.7	468	2	Q8RP89	Q8rp89 azoarcus sp
705	6	3.7	468	5	Q9XUB2	Q9xub2 caenorhabdi
706	6	3.7	470	2	Q9JMV6	Q9jmv6 bradyrhizob
707	6	3.7	470	10	Q9FPT2	Q9fpt2 arabidopsis
708	6	3.7	471	5	Q9Y1X6	Q9y1x6 ephydatia f
709	6	3.7	472	5	Q9V471	Q9v471 drosophila
710	6	3.7	473	2	Q9RP63	Q9rp63 escherichia
711	6	3.7	473	16	Q9PFC1	Q9pfc1 xylella fas
712	6	3.7	473	16	Q9KXW8	Q9kxw8 streptomyce
713	6	3.7	473	17	Q96XJ0	Q96xj0 sulfolobus
714	6	3.7	474	5	Q9Y1U0	Q9y1u0 giardia lam
715	6	3.7	474	5	Q8T4N3	Q8t4n3 rhipicephal
716	6	3.7	475	11	Q924A5	Q924a5 rattus norv
717	6	3.7	476	5	O61929	O61929 caenorhabdi
718	6	3.7	477	10	Q9LST0	Q9lst0 arabidopsis
719	6	3.7	477	16	Q9RD97	Q9rd97 streptomyce
720	6	3.7	478	2	Q9LBK7	Q9lbk7 gluconobact
721	6	3.7	478	4	Q9UIU5	Q9uiu5 homo sapien
722	6	3.7	478	4	Q9UBM8	Q9ubm8 homo sapien
723	6	3.7	478	12	Q91AX1	Q91ax1 human papil
724	6	3.7	481	16	Q8ZBM6	Q8zbm6 yersinia pe
725	6	3.7	482	2	Q9KHJ5	Q9khj5 streptomyce
726	6	3.7	482	2	P74912	P74912 thermobifid
727	6	3.7	482	5	Q9VLM3	Q9vlm3 drosophila
728	6	3.7	482	11	Q63207	Q63207 rattus norv
729	6	3.7	482	12	Q9WT28	Q9wt28 human herpe
730	6	3.7	482	16	Q8XQD2	Q8xqd2 ralstonia s
731	6	3.7	483	1	Q9P995	Q9p995 methanosarc
732	6	3.7	483	5	Q9VLM2	Q9vlm2 drosophila
733	6	3.7	483	11	Q8VD78	Q8vd78 mus musculu
734	6	3.7	483	11	Q8R135	Q8r135 mus musculu
735	6	3.7	484	11	Q99NB0	Q99nb0 mus musculu
736	6	3.7	484	11	Q9EP78	Q9ep78 mus musculu
737	6	3.7	484	12	Q9QJ33	Q9qj33 human herpe
738	6	3.7	484	16	Q9X8R5	Q9x8r5 streptomyce
739	6	3.7	485	16	Q92CN4	Q92cn4 listeria in
740	6	3.7	485	16	Q8Y7U7	Q8y7u7 listeria mo
741	6	3.7	485	16	Q8X5K8	Q8x5k8 escherichia
742	6	3.7	486	5	Q9VLM4	Q9vlm4 drosophila
743	6	3.7	487	10	O80556	O80556 arabidopsis
744	6	3.7	487	10	O82736	O82736 arabidopsis
745	6	3.7	488	10	Q8S133	Q8s133 oryza sativ
746	6	3.7	490	16	Q8XPT5	Q8xpt5 ralstonia s
747	6	3.7	491	5	Q9NGK1	Q9ngk1 entamoeba i
748	6	3.7	491	10	Q8W5D1	Q8w5d1 oryza sativ
749	6	3.7	491	16	P71906	P71906 mycobacteri
750	6	3.7	492	17	Q8TZ09	Q8tz09 methanopyru
751	6	3.7	493	12	Q89281	Q89281 yellow feve
752	6	3.7	493	12	Q89286	Q89286 yellow feve
753	6	3.7	493	16	Q9HVJ4	Q9hvj4 pseudomonas
754	6	3.7	495	8	O63853	O63853 sarcophyton
755	6	3.7	495	10	P98183	P98183 catharanthu
756	6	3.7	496	16	O53631	O53631 mycobacteri
757	6	3.7	498	11	Q9D001	Q9d001 mus musculu
758	6	3.7	498	16	Q98KQ0	Q98kq0 rhizobium l
759	6	3.7	499	16	Q9RYR7	Q9ryr7 deinococcus
760	6	3.7	500	5	Q9U2J0	Q9u2j0 caenorhabdi
761	6	3.7	500	5	Q95RG0	Q95rg0 drosophila
762	6	3.7	500	16	O83963	O83963 treponema p
763	6	3.7	500	16	Q989D1	Q989d1 rhizobium l
764	6	3.7	501	11	Q9QXF7	Q9qxf7 rattus norv
765	6	3.7	501	16	Q9RVX7	Q9rvx7 deinococcus
766	6	3.7	504	10	Q9SB41	Q9sb41 arabidopsis
767	6	3.7	507	12	Q9JGK2	Q9jgk2 soil-borne
768	6	3.7	507	12	Q9JGJ6	Q9jgj6 sorghum chl
769	6	3.7	508	16	P73757	P73757 synechocyst
770	6	3.7	509	10	Q9C9E6	Q9c9e6 arabidopsis
771	6	3.7	509	11	Q9QWI5	Q9qwi5 rattus norv
772	6	3.7	509	11	P97710	P97710 rattus norv
773	6	3.7	510	16	Q9PNY2	Q9pny2 campylobact
774	6	3.7	511	16	Q8VKT2	Q8vkt2 mycobacteri
775	6	3.7	511	16	Q8R949	Q8r949 thermoanaer
776	6	3.7	512	8	Q9MGD9	Q9mgd9 penaeus mon
777	6	3.7	512	12	Q69059	Q69059 human herpe
778	6	3.7	512	16	Q8YGF5	Q8ygf5 brucella me
779	6	3.7	513	8	Q9XMI0	Q9xmi0 farfantepen
780	6	3.7	513	8	Q94ST3	Q94st3 danaceticht
781	6	3.7	513	16	Q9AAJ4	Q9aaj4 caulobacter
782	6	3.7	514	10	Q9LLL6	Q91116 chlamydomon
783	6	3.7	514	10	Q94IP9	Q94ip9 lycopersico
784	6	3.7	514	16	P71588	P71588 mycobacteri
785	6	3.7	516	3	Q9Y782	Q9y782 coprinus ci
786	6	3.7	516	5	Q9VST9	Q9vst9 drosophila
787	6	3.7	516	10	Q945G6	Q945g6 prunus sero
788	6	3.7	516	17	Q979N1	Q979n1 thermoplasm
789	6	3.7	519	17	Q8TPG3	Q8tpg3 methanosarc
790	6	3.7	520	8	Q954T8	Q954t8 fenneropena
791	6	3.7	520	8	Q954T7	Q954t7 fenneropena
792	6	3.7	521	3	Q96TR4	Q96tr4 pleurotus o
793	6	3.7	522	2	O50271	O50271 agrobacteri
794	6	3.7	523	13	Q90YK5	Q90yk5 gallus gall
795	6	3.7	523	16	Q9KY50	Q9ky50 streptomyce
796	6	3.7	525	16	Q9KB26	Q9kb26 bacillus ha
797	6	3.7	525	17	Q9HMF2	Q9hmf2 halobacteri
798	6	3.7	527	12	Q9PYQ5	Q9pyq5 xestia c-ni
799	6	3.7	529	5	Q9V386	Q9v386 drosophila
800	6	3.7	530	10	Q9SQG7	Q9sqg7 lycopersico
801	6	3.7	530	10	Q9SQG6	Q9sqg6 lycopersico
802	6	3.7	530	10	O04431	O04431 nicotiana p
803	6	3.7	530	10	Q942S8	Q942s8 oryza sativ
804	6	3.7	531	5	Q19344	Q19344 caenorhabdi
805	6	3.7	531	10	Q43014	Q43014 prunus aviu
806	6	3.7	531	10	Q9FLU8	Q9flu8 arabidopsis
807	6	3.7	531	10	Q93X79	Q93x79 lycopersico
808	6	3.7	531	10	Q9ZW74	Q9zw74 arabidopsis
809	6	3.7	531	10	Q8RWR0	Q8rwr0 arabidopsis
810	6	3.7	533	16	Q8VKR5	Q8vkr5 mycobacteri
811	6	3.7	533	16	Q9CGU2	Q9cgu2 lactococcus
812	6	3.7	534	10	Q9FLU9	Q9flu9 arabidopsis
813	6	3.7	535	10	Q8RVG5	Q8rvg5 daucus caro
814	6	3.7	537	17	Q97BT4	Q97bt4 thermoplasm
815	6	3.7	539	3	Q12412	Q12412 saccharomyc
816	6	3.7	541	4	Q13478	Q13478 homo sapien
817	6	3.7	541	16	Q9RJ71	Q9rj71 streptomyce
818	6	3.7	543	2	Q9AJP4	Q9ajp4 streptomyce
819	6	3.7	544	4	Q13652	Q13652 homo sapien

820	6	3.7	544	10	Q9M5X4
821	6	3.7	545	17	Q980U2
822	6	3.7	548	16	Q9KU05
823	6	3.7	548	16	Q92NZ9
824	6	3.7	550	10	Q9LLM4
825	6	3.7	552	12	O72728
826	6	3.7	555	5	O18057
827	6	3.7	558	13	Q9PUB5
828	6	3.7	559	16	Q92ND1
829	6	3.7	562	9	Q37916
830	6	3.7	565	5	Q95V76
831	6	3.7	567	17	Q8TUM3
832	6	3.7	568	2	Q47466
833	6	3.7	571	3	Q9UU63
834	6	3.7	571	12	Q83848
835	6	3.7	572	16	Q8ZA26
836	6	3.7	573	5	Q22842
837	6	3.7	573	5	Q9VR24
838	6	3.7	573	12	Q83602
839	6	3.7	574	10	Q9SJQ2
840	6	3.7	574	12	Q83600
841	6	3.7	575	16	Q9I642
842	6	3.7	577	10	Q9XEW1
843	6	3.7	577	16	Q9K641
844	6	3.7	578	16	Q9HT06
845	6	3.7	580	10	Q9SIY8
846	6	3.7	580	10	Q8S8G1
847	6	3.7	581	16	Q98HN2
848	6	3.7	582	12	Q83345
849	6	3.7	582	12	Q83346
850	6	3.7	582	12	Q83347
851	6	3.7	582	12	Q83348
852	6	3.7	582	12	Q83650
853	6	3.7	582	12	Q9WAF0
854	6	3.7	582	12	Q9WAF3
855	6	3.7	582	12	Q9WAF5
856	6	3.7	582	12	Q9WAF8
857	6	3.7	582	12	Q9WAG1
858	6	3.7	582	12	Q9WAG4
859	6	3.7	582	12	Q9WAG6
860	6	3.7	582	12	Q9WAG8
861	6	3.7	582	12	Q9WAH0
862	6	3.7	582	12	Q9WAH2
863	6	3.7	582	12	Q9WAH5
864	6	3.7	582	12	Q9WAH7
865	6	3.7	582	12	Q9WAH9
866	6	3.7	582	12	Q9WAI1
867	6	3.7	582	12	Q9WAI3
868	6	3.7	582	12	Q9DQA1
869	6	3.7	582	12	Q9EAE4
870	6	3.7	582	12	Q9EAE3
871	6	3.7	582	12	Q9EAE2
872	6	3.7	582	12	Q9EAE1
873	6	3.7	582	12	Q9ICS5
874	6	3.7	582	12	Q83615
875	6	3.7	582	12	Q83652
876	6	3.7	582	12	Q83651
877	6	3.7	582	12	Q9WIX0
878	6	3.7	582	12	Q83649
879	6	3.7	582	12	Q9J4L1
880	6	3.7	582	12	Q91N69
881	6	3.7	582	12	Q91N68
882	6	3.7	582	12	Q91N64
883	6	3.7	582	12	Q8V299
884	6	3.7	582	12	Q8V298
885	6	3.7	582	12	Q8V297
886	6	3.7	582	12	Q8V296
887	6	3.7	582	12	Q8UYE4
888	6	3.7	582	12	Q89608
889	6	3.7	582	12	Q8QY68
890	6	3.7	582	12	Q8QV65
891	6	3.7	584	10	Q9SPU1
892	6	3.7	585	5	Q961E4

Q9m5x4 prunus sero
Q980u2 sulfolobus
Q9ku05 vibrio chol
Q92nz9 rhizobium m
Q9llm4 eucalyptus
O72728 cucurbit ye
O18057 caenorhabdi
Q9pub5 brachydanio
Q92ndl rhizobium m
Q37916 bacteriopha
Q95v76 leishmania
Q8tum3 methanosarc
Q47466 erwinia car
Q9uu63 schizosacch
Q83848 newcastle d
Q8za26 yersinia pe
Q22842 caenorhabdi
Q9vr24 drosophila
Q83602 mumps virus
Q9sjq2 arabidopsis
Q83600 mumps virus
Q9i642 pseudomonas
Q9xew1 arabidopsis
Q9k641 bacillus ha
Q9ht06 pseudomonas
Q9siy8 arabidopsis
Q8s8g1 arabidopsis
Q98hn2 rhizobium l
Q83345 mumps virus
Q83346 mumps virus
Q83347 mumps virus
Q83348 mumps virus
Q83650 mumps virus
Q9waf0 mumps virus
Q9waf3 mumps virus
Q9waf5 mumps virus
Q9waf8 mumps virus
Q9wag1 mumps virus
Q9wag4 mumps virus
Q9wag6 mumps virus
Q9wag8 mumps virus
Q9wah0 mumps virus
Q9wah2 mumps virus
Q9wah5 mumps virus
Q9wah7 mumps virus
Q9wah9 mumps virus
Q9wai1 mumps virus
Q9wai3 mumps virus
Q9dqa1 mumps virus
Q9eae4 mumps virus
Q9eae3 mumps virus
Q9eae2 mumps virus
Q9eae1 mumps virus
Q9ics5 mumps virus
Q83615 mumps virus
Q83652 mumps virus
Q83651 mumps virus
Q9wix0 mumps virus
Q83649 mumps virus
Q9j411 mumps virus
Q91n69 mumps virus
Q91n68 mumps virus
Q91n64 mumps virus
Q8v299 mumps virus
Q8v298 mumps virus
Q8v297 mumps virus
Q8v296 mumps virus
Q8uye4 mumps virus
Q89608 mumps virus
Q8qy68 mumps virus
Q8qv65 mumps virus
Q9spu1 oryza sativ
Q961e4 drosophila

893	6	3.7	585	5	Q8SSC7
894	6	3.7	586	5	Q9VQG9
895	6	3.7	586	11	Q9DCM7
896	6	3.7	586	11	Q8VDF8
897	6	3.7	587	4	Q96BF6
898	6	3.7	587	16	Q8U754
899	6	3.7	588	2	Q93L02
900	6	3.7	589	4	Q96AD6
901	6	3.7	590	11	O70292
902	6	3.7	590	11	O70297
903	6	3.7	590	11	Q8VEB1
904	6	3.7	592	9	Q9AZY8
905	6	3.7	592	16	Q9JXF1
906	6	3.7	592	16	Q9JWI7
907	6	3.7	592	16	Q9CGQ9
908	6	3.7	594	5	O16312
909	6	3.7	596	3	Q9URR8
910	6	3.7	596	5	Q93896
911	6	3.7	597	10	Q9LQ68
912	6	3.7	597	11	O35727
913	6	3.7	600	16	Q9RV23
914	6	3.7	606	5	Q9BLG5
915	6	3.7	606	5	Q18869
916	6	3.7	607	5	Q9TW52
917	6	3.7	607	16	P73606
918	6	3.7	610	5	Q9XVM2
919	6	3.7	610	16	Q9KRY4
920	6	3.7	612	5	Q9U003
921	6	3.7	612	11	Q9Z2N3
922	6	3.7	613	5	Q9VJ39
923	6	3.7	614	5	O45350
924	6	3.7	616	6	O97507
925	6	3.7	623	5	Q22976
926	6	3.7	623	5	Q8T902
927	6	3.7	624	13	Q90WU2
928	6	3.7	625	16	Q8R9T6
929	6	3.7	627	10	Q9SRP2
930	6	3.7	629	10	Q94D08
931	6	3.7	632	3	Q8X1F1
932	6	3.7	634	16	Q8XZP6
933	6	3.7	635	12	Q91MS8
934	6	3.7	635	12	Q8V3L4
935	6	3.7	636	5	Q21126
936	6	3.7	636	12	Q91TS2
937	6	3.7	637	12	Q8V4Y2
938	6	3.7	637	12	Q8V2R3
939	6	3.7	637	12	Q8QMV4
940	6	3.7	639	3	Q12537
941	6	3.7	643	16	Q9KMQ2
942	6	3.7	644	16	Q9RJK2
943	6	3.7	646	2	O51813
944	6	3.7	647	2	O86135
945	6	3.7	647	2	Q8VV97
946	6	3.7	650	5	Q24458
947	6	3.7	652	5	O96418
948	6	3.7	653	16	O06487
949	6	3.7	656	5	Q95YZ4
950	6	3.7	656	16	Q92ET6
951	6	3.7	660	1	Q9HHC8
952	6	3.7	661	5	O77164
953	6	3.7	661	16	Q98MM3
954	6	3.7	662	16	Q8U9L8
955	6	3.7	664	3	Q9UVU2
956	6	3.7	667	3	Q12101
957	6	3.7	669	5	Q9VM21
958	6	3.7	670	10	Q39717
959	6	3.7	671	5	P81538
960	6	3.7	671	16	Q98DL4
961	6	3.7	672	2	Q45343
962	6	3.7	679	10	Q9LPY5
963	6	3.7	681	3	O60157
964	6	3.7	684	5	Q961A0
965	6	3.7	685	17	O29057

Q8ssc7 encephalito
Q9vqg9 drosophila
Q9dcm7 mus musculu
Q8vdf8 mus musculu
Q96bf6 homo sapien
Q8u754 agrobacteri
Q93102 burkholderi
Q96ad6 homo sapien
O70292 mus musculu
O70297 mus musculu
Q8veb1 mus musculu
Q9azy8 bacteriopha
Q9jxf1 neisseria m
Q9jwi7 neisseria m
Q9cgq9 lactococcus
O16312 caenorhabdi
Q9urr8 penicillium
Q93896 caenorhabdi
Q9lq68 arabidopsis
O35727 mus musculu
Q9rv23 deinococcus
Q9blg5 ciona savig
Q18869 caenorhabdi
Q9tw52 caenorhabdi
P73606 synechocyst
Q9xvm2 caenorhabdi
Q9kry4 vibrio chol
Q9u003 caenorhabdi
Q9z2n3 mus musculu
Q9vj39 drosophila
O45350 caenorhabdi
O97507 sus scrofa
Q22976 caenorhabdi
Q8t902 drosophila
Q90wu2 fundulus he
Q8r9t6 thermoanaer
Q9srp2 arabidopsis
Q94d08 oryza sativ
Q8x1f1 cryptococcu
Q8xzp6 ralstonia s
Q91ms8 lumpy skin
Q8v3l4 swinepox vi
Q21126 caenorhabdi
Q91ts2 tupaia herp
Q8v4y2 monkeypox v
Q8v2r3 camelpox vi
Q8qmv4 cowpox viru
Q12537 aspergillus
Q9kmq2 vibrio chol
Q9rjk2 streptomyce
O51813 phormidium
O86135 bordetella
Q8vv97 bordetella
Q24458 drosophila
O96418 riptortus c
O06487 bacillus su
Q95yz4 leishmania
Q92et6 listeria in
Q9hhc8 thermococcu
O77164 entamoeba h
Q98mm3 rhizobium l
Q8u918 agrobacteri
Q9uvu2 pichia meth
Q12101 saccharomyc
Q9vm21 drosophila
Q39717 nicotiana t
P81538 drosophila
Q98dl4 rhizobium l
Q45343 bordetella
Q9lpy5 arabidopsis
O60157 schizosacch
Q961a0 drosophila
O29057 archaeoglob

966	6	3.7	693	3	Q9C4A2	Q9c4a2 tricholoma
967	6	3.7	693	12	Q69090	Q69090 human herpe
968	6	3.7	695	5	Q9W4C4	Q9w4c4 drosophila
969	6	3.7	695	16	Q8XR06	Q8xr06 ralstonia s
970	6	3.7	696	10	Q94DZ6	Q94dz6 oryza sativ
971	6	3.7	696	17	Q8ZTR8	Q8ztr8 pyrobaculum
972	6	3.7	698	17	O26996	O26996 methanobact
973	6	3.7	700	3	Q8TG00	Q8tg00 aspergillus
974	6	3.7	702	3	O74312	O74312 schizosacch
975	6	3.7	710	12	Q8V4W8	Q8v4w8 monkeypox v
976	6	3.7	710	12	Q8V2P8	Q8v2p8 camelpox vi
977	6	3.7	710	12	Q8QMU3	Q8qmu3 cowpox viru
978	6	3.7	711	12	Q9DHL5	Q9dhl5 yaba-like d
979	6	3.7	711	12	Q8V3K1	Q8v3k1 swinepox vi
980	6	3.7	714	12	Q91MR5	Q91mr5 lumpy skin
981	6	3.7	715	10	Q8W1S4	Q8w1s4 brassica na
982	6	3.7	716	13	Q91691	Q91691 xenopus lae
983	6	3.7	717	13	P70006	P70006 xenopus lae
984	6	3.7	718	16	Q9PQK8	Q9pqk8 ureaplasma
985	6	3.7	719	16	Q9AB85	Q9ab85 caulobacter
986	6	3.7	720	3	O14367	O14367 schizosacch
987	6	3.7	720	5	Q9GRN6	Q9grn6 leishmania
988	6	3.7	725	2	Q52978	Q52978 rhizobium m
989	6	3.7	725	5	Q9U1E0	Q9u1e0 leishmania
990	6	3.7	726	2	Q9AJB3	Q9ajb3 ruminococcu
991	6	3.7	727	5	Q19327	Q19327 caenorhabdi
992	6	3.7	727	10	Q9T0B4	Q9t0b4 arabidopsis
993	6	3.7	727	10	Q8SA09	Q8sa09 oryza sativ
994	6	3.7	730	5	Q95NF2	Q95nf2 drosophila
995	6	3.7	730	10	O65736	O65736 cicer ariet
996	6	3.7	732	17	Q979T3	Q979t3 thermoplasm
997	6	3.7	734	5	Q9VLT9	Q9vlt9 drosophila
998	6	3.7	734	16	Q9HT76	Q9ht76 pseudomonas
999	6	3.7	737	10	Q9M2Y4	Q9m2y4 arabidopsis
1000	6	3.7	737	10	Q8S6T7	Q8s6t7 oryza sativ

ALIGNMENTS

```
RESULT 1
Q9F5R9
ID   Q9F5R9      PRELIMINARY;      PRT;   166 AA.
AC   Q9F5R9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RA   Harris S.L., Spears P.A., Havell E.A., Hamrick T.S., Horton J.R.,
RA   Orndorff P.E.;
RT   "Isolation and characterization of Escherichia coli type 1 pili
RT   mutants that have altered binding specificities.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF306537; AAG30927.1; -.
SQ   SEQUENCE   166 AA;  17826 MW;  2E64EEA80AE286DC CRC64;

  Query Match             57.1%;  Score 92;  DB 2;  Length 166;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-88;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 2
Q9S497
ID   Q9S497      PRELIMINARY;      PRT;   300 AA.
AC   Q9S497;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=20327582; PubMed=10869080;
RA   Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA   Russell P.W., Orndorff P.E.;
RT   "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT   mutants and identification of a novel binding phenotype.";
RL   J. Bacteriol. 182:4012-4021(2000).
DR   EMBL; AF154925; AAD44319.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT      19      19       D -> A.
SQ   SEQUENCE   300 AA;  31517 MW;  D29340A1575A2358 CRC64;

  Query Match             57.1%;  Score 92;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-88;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 3
Q9S496
ID   Q9S496      PRELIMINARY;      PRT;   300 AA.
AC   Q9S496;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=20327582; PubMed=10869080;
RA   Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA   Russell P.W., Orndorff P.E.;
RT   "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT   mutants and identification of a novel binding phenotype.";
RL   J. Bacteriol. 182:4012-4021(2000).
DR   EMBL; AF154926; AAD44320.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT      21      21       L -> S.
SQ   SEQUENCE   300 AA;  31499 MW;  3ADF97A85F8CE478 CRC64;
```

```
  Query Match             57.1%;  Score 92;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-88;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 4
Q9F5S0
ID   Q9F5S0      PRELIMINARY;      PRT;   300 AA.
AC   Q9F5S0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=21289098; PubMed=11395476;
RA   Harris S.L., Spears P.A., Havell E.A., Hamrick T.S., Horton J.R.,
RA   Orndorff P.E.;
RT   "Characterization of Escherichia coli Type 1 Pilus Mutants with
RT   Altered Binding Specificities.";
RL   J. Bacteriol. 183:4099-4102(2001).
DR   EMBL; AF306536; AAG30926.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT     163     163       S -> F.
SQ   SEQUENCE   300 AA;  31413 MW;  922904051758746D CRC64;

  Query Match             57.1%;  Score 92;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-88;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 5
Q8XBA6
ID   Q8XBA6      PRELIMINARY;      PRT;   300 AA.
AC   Q8XBA6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Minor fimbrial subunit, D-mannose specific adhesin.
GN   FIMH OR Z5918 OR ECS5279.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005662; AAG59502.1; -.
DR   EMBL; AP002569; BAB38702.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
KW   Complete proteome.
SQ   SEQUENCE   300 AA;  31459 MW;  68EFE713A3D372CB CRC64;

  Query Match             55.9%;  Score 90;  DB 16;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-86;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db     109 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 6
Q9S495
ID   Q9S495      PRELIMINARY;      PRT;   300 AA.
AC   Q9S495;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=20327582; PubMed=10869080;
RA   Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA   Russell P.W., Orndorff P.E.;
RT   "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT   mutants and identification of a novel binding phenotype.";
RL   J. Bacteriol. 182:4012-4021(2000).
DR   EMBL; AF154927; AAD44321.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT      65      65       Y -> C.
SQ   SEQUENCE   300 AA;  31533 MW;  48FFB2795A3CB23D CRC64;

  Query Match             45.3%;  Score 73;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-68;
  Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      20 HNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWP 79
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66 HNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWP 125
```

```
Qy      80 VALYLTPVSSAGG 92
           |||||||||||||
Db     126 VALYLTPVSSAGG 138


RESULT 7
Q9S494
ID   Q9S494      PRELIMINARY;      PRT;   300 AA.
AC   Q9S494;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=20327582; PubMed=10869080;
RA   Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA   Russell P.W., Orndorff P.E.;
RT   "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT   mutants and identification of a novel binding phenotype.";
RL   J. Bacteriol. 182:4012-4021(2000).
DR   EMBL; AF154928; AAD44322.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT      79     79       R -> L.
SQ   SEQUENCE   300 AA;  31516 MW;  B6073DF68366B86E CRC64;

  Query Match             36.6%;  Score 59;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-53;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      34 QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 8
Q9F5S1
ID   Q9F5S1      PRELIMINARY;      PRT;   300 AA.
AC   Q9F5S1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=21289098; PubMed=11395476;
RA   Harris S.L., Spears P.A., Havell E.A., Hamrick T.S., Horton J.R.,
RA   Orndorff P.E.;
RT   "Characterization of Escherichia coli Type 1 Pilus Mutants with
RT   Altered Binding Specificities.";
RL   J. Bacteriol. 183:4099-4102(2001).
DR   EMBL; AF306535; AAG30925.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
FT   VARIANT      79     79       P -> R.
SQ   SEQUENCE   300 AA;  31457 MW;  9F7B0A35FF51F938 CRC64;

  Query Match             36.6%;  Score 59;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-53;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      34 QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 9
Q9R5Y2
ID   Q9R5Y2      PRELIMINARY;      PRT;   300 AA.
AC   Q9R5Y2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   FIMH SUBUNIT=MANNOSE-sensitive type 1 fimbrial adhesin.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94131954; PubMed=7905476;
RA   Sokurenko E.V., Courtney H.S., Ohman D.E., Klemm P., Hasty D.L.;
RL   J. Bacteriol. 176:748-755(1994).
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   300 AA;  31552 MW;  E75328B7C366A319 CRC64;

  Query Match             36.6%;  Score 59;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-53;
  Matches   59;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      34 QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80 QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 10
Q9S6R1
ID   Q9S6R1      PRELIMINARY;      PRT;   300 AA.
AC   Q9S6R1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Type 1 fimbriae adhesin, precursor polypeptide.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=88038337; PubMed=2890081;
RA   Klemm P., Christiansen G.;
RT   "Three fim genes required for the regulation of length and mediation
RT   of adhesion of Escherichia coli type 1 fimbriae.";
RL   Mol. Gen. Genet. 208:439-445(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=96235989; PubMed=8636962;
RA   Marc D., Dho-Moulin M.;
RT   "Analysis of the fim cluster of an avian O2 strain of Escherichia
RT   coli: serogroup-specific sites within fimA and nucleotide sequence of
RT   fimI.";
RL   J. Med. Microbiol. 44:444-452(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=96125229; PubMed=8550444;
RA   Klemm P., Tong S., Nielsen H., Conway T.;
RT   "The gntP gene of Escherichia coli involved in gluconate uptake.";
RL   J. Bacteriol. 178:61-67(1996).
```

```
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G.III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT78;
RX   MEDLINE=98439336; PubMed=9766199;
RA   Marc D., Arne P., Bree A., Dho-Moulin M.;
RT   "Colonization ability and pathogenic properties of a fim- mutant of an
RT   avian strain of Escherichia coli.";
RL   Res. Microbiol. 149:473-485(1998).
DR   EMBL; AJ225176; CAA12423.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   300 AA;  31459 MW;  EA040CEDD2137FE8 CRC64;

  Query Match             34.8%;  Score 56;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-50;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 196


RESULT 11
O87634
ID   O87634      PRELIMINARY;      PRT;   300 AA.
AC   O87634;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IHE3034;
RA   Pouttu R., Korhonen T.K.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF089840; AAC35864.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   300 AA;  31429 MW;  D789457A5F5255AE CRC64;

  Query Match             34.8%;  Score 56;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-50;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 196


RESULT 12
Q9F6Z7
ID   Q9F6Z7      PRELIMINARY;      PRT;   300 AA.
AC   Q9F6Z7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Type 1 fimbrial adhesin subunit FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LF82;
RX   MEDLINE=21150472; PubMed=11251843;
RA   Boudeau J., Barnich N., Darfeuille-Michaud A.;
RT   "Type 1 pili-mediated adherence of Escherichia coli strain LF82
RT   isolated from Crohn's disease is involved in bacterial invasion of
RT   intestinal epithelial cells.";
RL   Mol. Microbiol. 39:1272-1284(2001).
DR   EMBL; AF288194; AAG24827.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   300 AA;  31445 MW;  BA040E9DD2137E43 CRC64;

  Query Match             34.8%;  Score 56;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-50;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 196


RESULT 13
Q9AP05
ID   Q9AP05      PRELIMINARY;      PRT;   303 AA.
AC   Q9AP05;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RS218;
RA   Elliott S.J., Kim K.S.;
RT   "Virulence factors of meningitic Escherichia coli.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF317710; AAG60686.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
SQ   SEQUENCE   303 AA;  31773 MW;  4F615EA5D5E6763C CRC64;

  Query Match             34.8%;  Score 56;  DB 2;  Length 303;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-50;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     144 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 199


RESULT 14
Q9XBV8
ID   Q9XBV8      PRELIMINARY;      PRT;   129 AA.
AC   Q9XBV8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FimH.
GN   FIMH.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=20327582; PubMed=10869080;
RA   Hamrick T.S., Harris S.L., Spears P.A., Havell E.A., Horton J.R.,
RA   Russell P.W., Orndorff P.E.;
RT   "Genetic characterization of Escherichia coli type 1 pilus adhesin
RT   mutants and identification of a novel binding phenotype.";
RL   J. Bacteriol. 182:4012-4021(2000).
DR   EMBL; AF154929; AAD44323.1; -.
SQ   SEQUENCE   129 AA;  14142 MW;  BDE1D4D0E11EC536 CRC64;

  Query Match           31.1%;  Score 50;  DB 2;  Length 129;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-44;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTV 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTV 96


RESULT 15
Q08858
ID   Q08858      PRELIMINARY;      PRT;   243 AA.
AC   Q08858;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   FIMH protein precursor.
GN   FIMH.
OS   Klebsiella pneumoniae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Klebsiella.
OX   NCBI_TaxID=573;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89155420; PubMed=2563996;
RA   Gerlach G.F., Clegg S., Allen B.L.;
RT   "Identification and characterization of the genes encoding the type 3
RT   and type 1 fimbrial adhesins of Klebsiella pneumoniae.";
RL   J. Bacteriol. 171:1262-1270(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Nichols N.N., Clegg S.;
RL   Submitted (AUG-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
CC       ADHESION OF TYPE 1 FIMBRIAE (BUT NOT NECESSARY FOR THE PRODUCTION
CC       OF FIMBRIAE).
DR   EMBL; L23111; AAA25063.1; -.
KW   Fimbria; Signal.
FT   SIGNAL       1     25       POTENTIAL.
FT   CHAIN       26    243       FIMH PROTEIN.
SQ   SEQUENCE   243 AA;  25977 MW;  C8394D7C44A98AE7 CRC64;

  Query Match           26.1%;  Score 42;  DB 2;  Length 243;
  Best Local Similarity  100.0%;  Pred. No. 1e-35;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLS 44
          |||||||||||||||||||||||||||||||||||||||||||
Db     51 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLS 92


RESULT 16
Q998F5
ID   Q998F5      PRELIMINARY;      PRT;   295 AA.
AC   Q998F5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Probable capsid protein.
OS   porcine cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
OX   NCBI_TaxID=109993;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OF-1;
RA   Rupasinghe V., Iwatsuki-Horimoto K., Tajima T., Sugii S., Horimoto T.;
RT   "Analysis of the porcine cytomegalovirus major capsid protein gene.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051069; BAB40661.1; -.
DR   InterPro; IPR002690; Herpes_V23.
DR   Pfam; PF01802; Herpes_V23; 1.
SQ   SEQUENCE   295 AA;  33392 MW;  953641F47D07E620 CRC64;

  Query Match            5.6%;  Score 9;  DB 12;  Length 295;
  Best Local Similarity  100.0%;  Pred. No. 0.57;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    141 RDVTVTLPD 149
          |||||||||
Db    172 RDVTVTLPD 180


RESULT 17
Q8XAX2
ID   Q8XAX2      PRELIMINARY;      PRT;   304 AA.
AC   Q8XAX2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative adhesin, similar to FimH protein.
GN   Z2206 OR ECS2107.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005354; AAG56266.1; -.
DR   EMBL; AP002557; BAB35530.1; -.
DR   InterPro; IPR000259; Fimbrial.
DR   Pfam; PF00419; Fimbrial; 1.
KW   Complete proteome.
SQ   SEQUENCE   304 AA;  32041 MW;  EED538023D95AFD5 CRC64;

  Query Match            5.6%;  Score 9;  DB 16;  Length 304;
  Best Local Similarity  100.0%;  Pred. No. 0.59;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 GQNLVVDLS 14
          |||||||||
Db     53 GQNLVVDLS 61
```

```
RESULT 18
Q9PT94
ID   Q9PT94       PRELIMINARY;      PRT;   550 AA.
AC   Q9PT94;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Winged helix nude.
GN   WHN.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20509988; PubMed=11054532;
RA   Schlake T., Schorpp M., Boehm T.;
RT   "Formation of regulator/target gene relationships during evolution.";
RL   Gene 256:29-34(2000).
DR   EMBL; AJ252024; CAB64948.1; -.
DR   HSSP; Q63245; 2HFH.
DR   InterPro; IPR001766; TF_Fork_head.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM00339; FH; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
SQ   SEQUENCE   550 AA;  60963 MW;  A23748F2022B3792 CRC64;

  Query Match             5.6%;  Score 9;  DB 13;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TPVSSAGGL 93
           |||||||||
Db     505 TPVSSAGGL 513


RESULT 19
Q9DG30
ID   Q9DG30       PRELIMINARY;      PRT;   550 AA.
AC   Q9DG30;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transcription factor Foxn4.
GN   FOXN4.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TUEBINGEN; TISSUE=THYMUS;
RA   Visel A., Willett C.E., Danilova N.P., Steiner L.A.;
RT   "Expression of winged-helix transcription factor foxn4 (whn) in
RT   zebrafish.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF198446; AAG27086.2; -.
DR   HSSP; Q63245; 2HFH.
DR   InterPro; IPR001766; TF_Fork_head.
DR   Pfam; PF00250; Fork_head; 2.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM00339; FH; 2.
DR   PROSITE; PS00658; FORK_HEAD_2; 2.
DR   PROSITE; PS50039; FORK_HEAD_3; 2.
SQ   SEQUENCE   550 AA;  60977 MW;  A8E3D8F2022B3797 CRC64;

  Query Match             5.6%;  Score 9;  DB 13;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TPVSSAGGL 93
           |||||||||
Db     505 TPVSSAGGL 513


RESULT 20
Q9RWQ4
ID   Q9RWQ4       PRELIMINARY;      PRT;   539 AA.
AC   Q9RWQ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Arginine utilization protein ROCB, putative.
GN   DR0612.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001919; AAF10189.1; -.
DR   TIGR; DR0612; -.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
KW   Complete proteome.
SQ   SEQUENCE   539 AA;  57561 MW;  946BFD6F7FDFF37B CRC64;

  Query Match             5.0%;  Score 8;  DB 16;  Length 539;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     137 DVSARDVT 144
           ||||||||
Db       4 DVSARDVT 11


RESULT 21
Q9H1Y7
ID   Q9H1Y7       PRELIMINARY;      PRT;   564 AA.
AC   Q9H1Y7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   DJ862P8.3 (Similar to MAP3K10 (Mitogen-activated protein kinase kinase
DE   kinase 10)) (Fragment).
GN   DJ862P8.3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hall R.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; AL133380; CAC17571.1; -.
DR   HSSP; P06241; 1SHF.
```

```
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Kinase; SH3 domain.
FT   NON_TER     564     564
SQ   SEQUENCE   564 AA;  62400 MW;  D4F91E14842E7EEB CRC64;

  Query Match             5.0%;  Score 8;  DB 4;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TPVSSAGG 92
           ||||||||
Db      12 TPVSSAGG 19


RESULT 22
Q8WWN2
ID   Q8WWN2      PRELIMINARY;      PRT;   570 AA.
AC   Q8WWN2;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Mixed lineage kinase 4alpha.
GN   MLK4ALPHA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RA   Kvasha S., Protopopov A., Rynditch A., Zabarovsky E., Kashuba V.;
RT   "MLK4, a new member of mixed lineage kinases.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ311797; CAC84639.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN_1.
DR   PROSITE; PS50002; SH3; 1.
KW   Kinase.
SQ   SEQUENCE   570 AA;  62922 MW;  34CE937FCBCC5F93 CRC64;

  Query Match             5.0%;  Score 8;  DB 4;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TPVSSAGG 92
           ||||||||
Db      12 TPVSSAGG 19


RESULT 23
Q80929
ID   Q80929      PRELIMINARY;      PRT;   599 AA.
AC   Q80929;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Replication protein E1.
GN   E1.
OS   Human papillomavirus type 50.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=40539;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Delius H.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Farmer A.D.;
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
CC       VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
CC       PROTEIN. THE E1-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
CC       CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
DR   EMBL; U31790; AAA79473.1; -.
DR   InterPro; IPR001177; Papillom_E1.
DR   Pfam; PF00519; E1; 1.
DR   Pfam; PF00524; E1_N; 1.
KW   Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
KW   Nuclear protein.
FT   NP_BIND     427     434       ATP (POTENTIAL).
SQ   SEQUENCE   599 AA;  68862 MW;  EA8D2ED3E8C63D99 CRC64;

  Query Match             5.0%;  Score 8;  DB 12;  Length 599;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     109 TNNYNSDD 116
           ||||||||
Db     381 TNNYNSDD 388


RESULT 24
Q9SHB7
ID   Q9SHB7      PRELIMINARY;      PRT;   860 AA.
AC   Q9SHB7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   At2g40720 protein.
GN   AT2G40720.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
```

```
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007660; AAD32807.1; -.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR002885; PPR.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF01535; PPR; 18.
DR   TIGRFAMs; TIGR00756; PPR; 12.
DR   PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.
DR   PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
SQ   SEQUENCE   860 AA;  95393 MW;  A79159C501D77976 CRC64;

  Query Match             5.0%;  Score 8;  DB 10;  Length 860;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     138 VSARDVTV 145
           ||||||||
Db     128 VSARDVTV 135


RESULT 25
Q8WWN1
ID   Q8WWN1      PRELIMINARY;      PRT;  1036 AA.
AC   Q8WWN1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Mixed lineage kinase 4beta.
GN   MLK4BETA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kvasha S., Protopopov A., Rynditch A., Zabarovsky E., Kashuba V.;
RT   "MLK4, a new member of mixed lineage kinases.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ311798; CAC84640.1; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00326; SH3; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN_1.
DR   PROSITE; PS50002; SH3; 1.
KW   Kinase.
SQ   SEQUENCE   1036 AA;  113809 MW;  B9C2ACF397C2CF37 CRC64;

  Query Match             5.0%;  Score 8;  DB 4;  Length 1036;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TPVSSAGG 92
           ||||||||
Db      12 TPVSSAGG 19


RESULT 26
O01505
ID   O01505      PRELIMINARY;      PRT;  1239 AA.
AC   O01505;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 135.4 kDa protein.
GN   C37A2.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Le T.T., Kemp K., Scheet P.;
RT   "The sequence of C. elegans cosmid C37A2.";
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U97194; AAB52447.2; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
KW   Hypothetical protein.
SQ   SEQUENCE   1239 AA;  135436 MW;  81F6C299FDC77D85 CRC64;

  Query Match             5.0%;  Score 8;  DB 5;  Length 1239;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     151 RGSVPIPL 158
           ||||||||
Db     960 RGSVPIPL 967


RESULT 27
Q9Z4N7
ID   Q9Z4N7      PRELIMINARY;      PRT;  1873 AA.
AC   Q9Z4N7;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Surface protein precursor.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Enterococcaceae; Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MMH594;
RX   MEDLINE=99081742; PubMed=9864215;
RA   Shankar V., Baghdayan A.S., Huycke M.M., Lindahl G., Gilmore M.S.;
RT   "Infection-derived Enterococcus faecalis strains are enriched in esp,
RT   a gene encoding a novel surface protein.";
RL   Infect. Immun. 67:193-200(1999).
DR   EMBL; AF034779; AAD09858.1; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
```

```
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR   TIGRFAMs; TIGR01168; YSIRK_signal; 1.
KW   Signal.
FT   SIGNAL        1     49       POTENTIAL.
FT   CHAIN        50   1873       SURFACE PROTEIN.
SQ   SEQUENCE   1873 AA;  202084 MW;  F609483DB9B80621 CRC64;

  Query Match             5.0%;  Score 8;  DB 2;  Length 1873;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    142 DVTVTLPD 149
          ||||||||
Db   1754 DVTVTLPD 1761


RESULT 28
Q9BJG6
ID   Q9BJG6      PRELIMINARY;      PRT;    63 AA.
AC   Q9BJG6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Variant surface protein (Fragment).
GN   VAR.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=209;
RA   Nogueira P.A., Wunderlich G., Tada M.S., Costa J.D.N., Menezes M.J.,
RA   Scherf A., Pereira da Silva L.H.;
RT   "Plasmodium falciparum: repertoire of expressed var genes and adhesion
RT   properties to endothelial receptors of clinical isolates from patients
RT   in Rondonia (Brazilian western Amazon region).";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF333312; AAK19583.1; -.
FT   NON_TER       1      1
FT   NON_TER      63     63
SQ   SEQUENCE   63 AA;  7108 MW;  ED958F8E1B8B85F7 CRC64;

  Query Match             4.3%;  Score 7;  DB 5;  Length 63;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    108 QTNNYNS 114
          |||||||
Db     13 QTNNYNS 19


RESULT 29
O75702
ID   O75702      PRELIMINARY;      PRT;    73 AA.
AC   O75702;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Protein-tyrosine-phosphatase, isoform 3 (EC 3.1.3.48).
GN   ACP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   Modesti A., Marzocchini R., Raugei G., Chiti A., Sereni A.,
RA   Magherini F., Ramponi G.;
RT   "Cloning, expression and characterisation of a new human low Mr
RT   phosphotyrosine protein phosphatase originating by alternative
RT   splicing.";
RL   FEBS Lett. 0:0-0(0).
DR   EMBL; Y16846; CAA76416.1; -.
DR   HSSP; P24666; 5PNT.
DR   InterPro; IPR000106; Low_mwt_PTPase.
DR   Pfam; PF01451; LMWPc; 1.
DR   SMART; SM00226; LMWPc; 1.
KW   Hydrolase.
SQ   SEQUENCE   73 AA;  7660 MW;  D261205427CBECF3 CRC64;

  Query Match             4.3%;  Score 7;  DB 4;  Length 73;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     98 GSLIAVL 104
          |||||||
Db     50 GSLIAVL 56


RESULT 30
Q9ND29
ID   Q9ND29      PRELIMINARY;      PRT;    74 AA.
AC   Q9ND29;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Variant surface protein (Fragment).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MT08;
RX   MEDLINE=20372615; PubMed=10910718;
RA   Kirchgatter K., Mosbach R., del Portillo H.A.;
RT   "Plasmodium falciparum: DBL-1 var sequence analysis in field isolates
RT   from central Brazil.";
RL   Exp. Parasitol. 95:154-157(2000).
DR   EMBL; AF172785; AAF89775.1; -.
FT   NON_TER       1      1
FT   NON_TER      74     74
SQ   SEQUENCE   74 AA;  8328 MW;  DFA8947BF49A2E61 CRC64;

  Query Match             4.3%;  Score 7;  DB 5;  Length 74;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    108 QTNNYNS 114
          |||||||
Db     16 QTNNYNS 22


RESULT 31
O31846
ID   O31846      PRELIMINARY;      PRT;    87 AA.
AC   O31846;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   YOZN protein.
GN   YOZN.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
```

```
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99114; CAB13819.1; -.
KW   Complete proteome.
SQ   SEQUENCE   87 AA;  9671 MW;  D36E195DA06C2C24 CRC64;

  Query Match             4.3%;  Score 7;  DB 16;  Length 87;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNL 9
           |||||||
Db      26 VNVGQNL 32


RESULT 32
Q9N3B4
ID   Q9N3B4      PRELIMINARY;      PRT;   107 AA.
AC   Q9N3B4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 11.9 kDa protein.
GN   Y54G2A.19.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wollam C., Zidanic M.;
RT   "The sequence of C. elegans cosmid Y54G2A.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC024817; AAF59579.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   107 AA;  11887 MW;  5A8456C4FB5B1D85 CRC64;

  Query Match             4.3%;  Score 7;  DB 5;  Length 107;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      69 VYNSRTD 75
           |||||||
Db      69 VYNSRTD 75


RESULT 33
Q9FE65
ID   Q9FE65      PRELIMINARY;      PRT;   119 AA.
AC   Q9FE65;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative 60S ribosomal protein L34 (60S ribosomal protein L34,
DE   putative) (T6C23.18) (F24J1.23/F24J1.23).
GN   T6C23.18 OR F24J1.23.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sakano H., Liu S.X., Pham P.K., Yamada K., Banh J., Etgu P., Lee J.M.,
RA   Toriumi M., Yu G., Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "Full Length cDNA of gene T6C23.18 (GI:6665554).";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA   Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA   Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT   "Arabidopsis thaliana chromosome 1 BAC F24J1 genomic sequence.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Brooks S., Chao Q., Chen H., Kim C., Ecker J.;
RT   "Full length cDNA sequences of Arabidopsis thaliana.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA   Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
RA   Carninici P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA   Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA   Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA   Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA   Theologis A.;
RT   "Full Length cDNA of gene T6C23.18 (GI:6665554).";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
```

```
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzerg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA   Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA   Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA   Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA   Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA   Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA   Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA   Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis cDNA clones.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RA   Kim C.J., Chen H., Cheuk R., Meyers M.C., Banh J., Bowser L.,
RA   Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA   Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA   Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA   Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA   Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
RA   Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis ORF clones.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF327531; AAG42912.1; -.
DR   EMBL; AC021046; AAG12705.1; -.
DR   EMBL; AF324703; AAG40054.1; -.
DR   EMBL; AF349526; AAK15573.1; -.
DR   EMBL; AC013289; AAG52537.1; -.
DR   EMBL; AY052720; AAK96624.1; -.
DR   EMBL; AF446885; AAL38618.1; -.
DR   InterPro; IPR001284; Ribosomal_L34E.
DR   Pfam; PF01199; Ribosomal_L34e; 1.
DR   PRINTS; PR01250; RIBOSOMALL34.
DR   ProDom; PD005148; Ribosomal_L34E; 1.
DR   PROSITE; PS01145; RIBOSOMAL_L34E; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   119 AA;  13650 MW;  9116C7E50F46E627 CRC64;

  Query Match             4.3%;  Score 7;  DB 10;  Length 119;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 AYGGVLS 44
           |||||||
Db      77 AYGGVLS 83


RESULT 34
Q9LJW6
ID   Q9LJW6      PRELIMINARY;      PRT;   120 AA.
AC   Q9LJW6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   60S ribosomal protein L34 (Putative 60S ribosomal protein L34).
GN   AT3G28900.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=20363099; PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
RA   Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
RA   Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA   Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA   Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA   Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA   Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA   Ecker J.R., Theologis A.;
RT   "Full Length cDNA of gene MLD15.7/AT3g28900 (GI:9294231).";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA   Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA   Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA   Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA   Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA   Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA   Davis R.W., Ecker J.R., Theologis A.;
RT   "Arabidopsis Open Reading Frame (ORF) Clones.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP000386; BAB02133.1; -.
DR   EMBL; AY045818; AAK76492.1; -.
DR   EMBL; AY091359; AAM14298.1; -.
DR   InterPro; IPR001284; Ribosomal_L34E.
DR   Pfam; PF01199; Ribosomal_L34e; 1.
DR   PRINTS; PR01250; RIBOSOMALL34.
DR   ProDom; PD005148; Ribosomal_L34E; 1.
DR   PROSITE; PS01145; RIBOSOMAL_L34E; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   120 AA;  13650 MW;  2C3D7E7631FBDF5A CRC64;

  Query Match             4.3%;  Score 7;  DB 10;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 AYGGVLS 44
           |||||||
Db      77 AYGGVLS 83


RESULT 35
Q9RXH5
ID   Q9RXH5      PRELIMINARY;      PRT;   128 AA.
AC   Q9RXH5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein DR0338.
GN   DR0338.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001894; AAF09928.1; -.
DR   TIGR; DR0338; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   128 AA;  14083 MW;  63AE3D9F10FEE928 CRC64;

  Query Match             4.3%;  Score 7;  DB 16;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      43 LSNFSGT 49
           |||||||
Db      23 LSNFSGT 29


RESULT 36
Q8U8J1
ID   Q8U8J1      PRELIMINARY;      PRT;   162 AA.
AC   Q8U8J1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Atu4100.
GN   ATU4100 OR AGR_L_1506.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009341; AAL44901.1; -.
DR   EMBL; AE008274; AAK89331.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   162 AA;  18483 MW;  6362912EABE190B0 CRC64;

  Query Match             4.3%;  Score 7;  DB 16;  Length 162;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 NVGQNLV 10
           |||||||
Db      56 NVGQNLV 62


RESULT 37
Q8VV80
ID   Q8VV80      PRELIMINARY;      PRT;   177 AA.
AC   Q8VV80;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   F0F1-ATPase subunit delta.
GN   ATPH.
OS   Colwellia maris (Vibrio sp. (strain ABE-1)).
OC   Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
OC   Colwellia.
OX   NCBI_TaxID=77524;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Takada Y., Takiya S.;
RT   "Colwellia maris atp operon, complete sequence.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB035129; BAB82481.1; -.
DR   InterPro; IPR000711; ATPsynt_OSCP.
DR   Pfam; PF00213; OSCP; 1.
DR   PRINTS; PR00125; ATPASEDELTA.
DR   PROSITE; PS00389; ATPASE_DELTA; UNKNOWN_1.
SQ   SEQUENCE   177 AA;  19131 MW;  1CF62614457A0A78 CRC64;

  Query Match             4.3%;  Score 7;  DB 2;  Length 177;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      92 GLVIKAG 98
           |||||||
Db     150 GLVIKAG 156


RESULT 38
Q9U7J9
ID   Q9U7J9      PRELIMINARY;      PRT;   179 AA.
AC   Q9U7J9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Erythrocyte membrane protein 1 SD105E (Fragment).
GN   VAR.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99404835; PubMed=10477185;
RA   Ward C.P., Clottey G.T., Dorris M., Ji D.D., Arnot D.E.;
RT   "Analysis of Plasmodium falciparum PfEMP-1/var genes suggests that
RT   recombination rearranges constrained sequences.";
RL   Mol. Biochem. Parasitol. 102:167-177(1999).
DR   EMBL; AF127288; AAD52768.1; -.
FT   NON_TER       1      1
FT   NON_TER     179    179
SQ   SEQUENCE   179 AA;  20577 MW;  F625DAA80FC75086 CRC64;

  Query Match             4.3%;  Score 7;  DB 5;  Length 179;
  Best Local Similarity   100.0%;  Pred. No. 46;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     108 QTNNYNS 114
           |||||||
Db       8 QTNNYNS 14
```

```
RESULT 39
Q9PDQ1
ID   Q9PDQ1      PRELIMINARY;      PRT;   188 AA.
AC   Q9PDQ1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Cytochrome B561.
GN   XF1328.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE003965; AAF84137.1; -.
KW   Complete proteome.
SQ   SEQUENCE   188 AA;  21209 MW;  1A22CAFBA0BAEA82 CRC64;

  Query Match             4.3%;  Score 7;  DB 16;  Length 188;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     100 LIAVLIL 106
           |||||||
Db      20 LIAVLIL 26


RESULT 40
Q9CHD6
ID   Q9CHD6      PRELIMINARY;      PRT;   210 AA.
AC   Q9CHD6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein yiaD.
GN   YIAD OR LL0796.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL1403;
RX   MEDLINE=21235186; PubMed=11337471;
RA   Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA   Weissenbach J., Ehrlich S.D., Sorokin A.;
RT   "The complete genome sequence of the lactic acid bacterium Lactococcus
RT   lactis ssp. lactis IL1403.";
RL   Genome Res. 11:731-753(2001).
DR   EMBL; AE006313; AAK04894.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   210 AA;  23075 MW;  00025205406FC480 CRC64;

  Query Match             4.3%;  Score 7;  DB 16;  Length 210;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      96 KAGSLIA 102
           |||||||
Db      12 KAGSLIA 18


RESULT 41
O05437
ID   O05437      PRELIMINARY;      PRT;   248 AA.
AC   O05437;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Rv3908.
GN   RV3908 OR MTCY15F10.03C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; Z94121; CAB08093.1; -.
DR   TubercuList; Rv3908; -.
DR   InterPro; IPR000086; NUDIX_hydrolase.
DR   Pfam; PF00293; NUDIX; 1.
DR   PRINTS; PR00502; NUDIXFAMILY.
DR   PROSITE; PS00893; NUDIX; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   248 AA;  27171 MW;  270A5193861EADD3 CRC64;

  Query Match             4.3%;  Score 7;  DB 16;  Length 248;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLVI 95
           |||||||
Db      66 SAGGLVI 72


RESULT 42
Q8VIR0
ID   Q8VIR0      PRELIMINARY;      PRT;   248 AA.
AC   Q8VIR0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
```

```
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   MutT/nudix family protein.
GN   MT4027.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE007193; AAK48392.1; -.
DR   TIGR; MT4027; -.
DR   InterPro; IPR000086; NUDIX_hydrolase.
DR   Pfam; PF00293; NUDIX; 1.
DR   PRINTS; PR00502; NUDIXFAMILY.
DR   PROSITE; PS00893; NUDIX; UNKNOWN_1.
SQ   SEQUENCE   248 AA;  27125 MW;  989C08F615D19120 CRC64;

  Query Match             4.3%;  Score 7;  DB 16;  Length 248;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLVI 95
           |||||||
Db      66 SAGGLVI 72


RESULT 43
Q50195
ID   Q50195      PRELIMINARY;      PRT;   251 AA.
AC   Q50195;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   L222-ORF6 (Hypothetical protein ML2698).
GN   ML2698.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97124199; PubMed=8969512;
RA   Fsihi H., De Rossi E., Salazar L., Cantoni R., Labo M., Riccardi G.,
RA   Takiff H.E., Eiglmeier K., Bergh S., Cole S.T.;
RT   "Gene arrangement and organization in a approximately 76 kb fragment
RT   encompassing the oriC region of the chromosome of Mycobacterium
RT   leprae.";
RL   Microbiology 142:3147-3161(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
DR   EMBL; L39923; AAB53126.1; -.
DR   EMBL; AL583926; CAC32230.1; -.
DR   Leproma; ML2698; -.
DR   InterPro; IPR000086; NUDIX_hydrolase.
DR   Pfam; PF00293; NUDIX; 1.
DR   PRINTS; PR00502; NUDIXFAMILY.
DR   PROSITE; PS00893; NUDIX; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   251 AA;  28083 MW;  86F4E04D734219F6 CRC64;

  Query Match             4.3%;  Score 7;  DB 16;  Length 251;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLVI 95
           |||||||
Db      69 SAGGLVI 75


RESULT 44
Q8R805
ID   Q8R805      PRELIMINARY;      PRT;   256 AA.
AC   Q8R805;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Dehydrogenases with different specificities (related to short-chain
DE   alcohol dehydrogenases).
GN   FABG5 OR TTE2234.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4T / JCM11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013167; AAM25384.1; -.
KW   Complete proteome.
SQ   SEQUENCE   256 AA;  28153 MW;  B08E48F177C61314 CRC64;

  Query Match             4.3%;  Score 7;  DB 16;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 GQNLVVD 12
           |||||||
Db     244 GQNLVVD 250


RESULT 45
Q9ZVF6
ID   Q9ZVF6      PRELIMINARY;      PRT;   283 AA.
AC   Q9ZVF6; Q93Z95;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Expressed protein (At2g01490/F2I9.11).
GN   AT2G01490.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
```

RA Shen M., Ronning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Town C.D., Kaul S.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AC005560; AAC67325.2; -.
DR EMBL; AY057710; AAL15340.1; -.
SQ SEQUENCE 283 AA; 32028 MW; 1C4520E012ABD1C7 CRC64;

Query Match 4.3%; Score 7; DB 10; Length 283;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 96 KAGSLIA 102
|||||||
Db 225 KAGSLIA 231

Search completed: November 28, 2002, 19:05:33
Job time : 99 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 19:00:51 ; Search time 19 Seconds
(without alignments)
814.613 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 161
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	92	57.1	300	2	S56545	fimbrial protein f
2	90	55.9	300	2	G91288	hypothetical prote
3	90	55.9	300	2	B86130	hypothetical prote
4	42	26.1	302	2	A32801	fimbrial adhesin p
5	9	5.6	304	2	A64904	probable fimbrial
6	9	5.6	304	2	C90892	probable adhesin [
7	9	5.6	304	2	F85725	probable adhesin,
8	8	5.0	539	2	F75497	probable arginine
9	8	5.0	860	2	A84833	hypothetical prote
10	8	5.0	1234	2	T30160	hypothetical prote
11	8	5.0	1873	2	T30944	surface protein pr
12	7	4.3	87	2	B69932	hypothetical prote
13	7	4.3	119	2	H96717	probable 60S ribos
14	7	4.3	120	2	S48027	ribosomal protein
15	7	4.3	120	2	S60476	ribosomal protein
16	7	4.3	120	2	F86395	60s ribosomal prot
17	7	4.3	128	2	B75531	hypothetical prote
18	7	4.3	132	1	MPHU2	myelin P2 protein
19	7	4.3	162	2	A98226	hypothetical prote
20	7	4.3	162	2	AG3060	conserved hypothet
21	7	4.3	188	2	B82696	cytochrome B561 XF
22	7	4.3	210	2	D86724	hypothetical prote
23	7	4.3	248	2	E70600	hypothetical prote
24	7	4.3	251	2	H87246	conserved hypothet
25	7	4.3	291	2	F69959	hypothetical prote
26	7	4.3	292	2	S24169	mucin - rat
27	7	4.3	298	2	E97096	Zn-binding lipopro
28	7	4.3	304	2	H75081	phosphate abc tran
29	7	4.3	311	2	T32776	hypothetical prote
30	7	4.3	338	2	T38762	hypothetical prote
31	7	4.3	383	2	AD0183	probable exported
32	7	4.3	406	2	AG2704	MFS permease [impo
33	7	4.3	414	2	G97486	probable mfS trans
34	7	4.3	443	2	B86849	polysacharide bios
35	7	4.3	498	2	F98149	hypothetical prote
36	7	4.3	533	2	AE3138	hypothetical prote
37	7	4.3	590	2	F75501	sensor histidine k
38	7	4.3	629	2	S29685	retroviral recepto
39	7	4.3	650	2	G84334	threonyl-tRNA synt
40	7	4.3	687	2	S74621	carbon dioxide con
41	7	4.3	741	2	D81798	lactoferrin-bindin
42	7	4.3	761	2	G90728	probable enzyme [i
43	7	4.3	761	2	H85579	probable enzyme yb
44	7	4.3	761	2	C64813	ybhJ protein - Esc
45	7	4.3	790	2	T30081	hypothetical prote
46	7	4.3	839	2	D97013	probably secreted
47	7	4.3	1001	2	C86181	hypothetical prote
48	7	4.3	1004	2	T00795	26S proteasome reg
49	7	4.3	1040	2	B64974	hypothetical prote
50	7	4.3	1040	2	C90989	hypothetical prote
51	7	4.3	1040	2	F85834	hypothetical prote
52	7	4.3	1040	2	AD0771	probable RND-famil
53	7	4.3	1043	2	H83329	probable RND efflu
54	7	4.3	1052	2	AI0346	AcrB/AcrD/AcrF fam
55	7	4.3	1939	2	D97316	probable S-layer p
56	6	3.7	14	2	S57574	T cell receptor V-
57	6	3.7	29	2	S46929	teg169 protein - m
58	6	3.7	43	2	S53456	superoxide dismuta
59	6	3.7	60	2	D82806	hypothetical prote
60	6	3.7	68	1	WMBPUB	gene o protein - p
61	6	3.7	68	2	JQ2005	hypothetical 7.5K
62	6	3.7	73	2	T07913	probable aspartic
63	6	3.7	90	2	T18068	hypothetical prote
64	6	3.7	91	2	F69069	hypothetical prote
65	6	3.7	94	2	G69799	conserved hypothet
66	6	3.7	100	2	F72751	hypothetical prote
67	6	3.7	111	2	S51353	cytochrome-c oxida
68	6	3.7	111	2	E71063	hypothetical prote
69	6	3.7	111	2	E75084	hypothetical prote
70	6	3.7	111	2	A64917	probable membrane
71	6	3.7	111	2	A90918	hypothetical prote
72	6	3.7	111	2	F85766	hypothetical prote
73	6	3.7	111	2	AB0690	probable membrane
74	6	3.7	115	2	B86885	50S ribosomal prot
75	6	3.7	118	2	H81666	conserved hypothet
76	6	3.7	119	2	D97043	hypothetical prote
77	6	3.7	122	2	C37281	replication factor
78	6	3.7	126	2	I60079	opsin, middle-wave
79	6	3.7	128	2	A35690	mucin 3 (clone SIB
80	6	3.7	130	2	G81254	30S ribosomal prot
81	6	3.7	131	2	C97282	ribosomal protein
82	6	3.7	131	2	JC4767	E3 gene encoding h
83	6	3.7	133	2	JC2003	NADH ubiquinone ox
84	6	3.7	134	2	C72341	propionyl-CoA carb
85	6	3.7	136	1	PN0501	phosphoribosylanth
86	6	3.7	137	1	F64961	hypothetical prote
87	6	3.7	137	2	C85815	hypothetical prote
88	6	3.7	137	2	D90967	hypothetical prote
89	6	3.7	139	2	S77060	transposase sll066
90	6	3.7	144	2	A96580	hypothetical prote
91	6	3.7	145	2	B87560	hypothetical prote
92	6	3.7	147	2	AB1791	hypothetical prote
93	6	3.7	147	2	AC1415	hypothetical prote
94	6	3.7	149	2	S60474	histone H2A - gard
95	6	3.7	150	2	JQ1183	histone H2A - gard
96	6	3.7	153	2	T05637	hypothetical prote
97	6	3.7	154	2	D97415	hypothetical prote
98	6	3.7	155	2	T22815	hypothetical prote
99	6	3.7	156	2	T23109	hypothetical prote
100	6	3.7	157	2	E83629	hypothetical prote
101	6	3.7	159	2	S42986	probable signal pe
102	6	3.7	160	2	E95882	hypothetical trans

103	6	3.7	161	2	F82637	conserved hypothet
104	6	3.7	163	2	G72707	hypothetical prote
105	6	3.7	164	2	A81325	probable signal-tr
106	6	3.7	165	2	S43898	mcrD protein - Met
107	6	3.7	165	2	D69493	hypothetical prote
108	6	3.7	167	2	A59091	hypothetical prote
109	6	3.7	168	2	C37840	hypothetical prote
110	6	3.7	178	2	B71824	hypothetical prote
111	6	3.7	179	2	B69695	ribosomal protein
112	6	3.7	179	2	C81732	peptidyl-tRNA hydr
113	6	3.7	179	2	AB1974	hypothetical prote
114	6	3.7	185	2	AG2026	hypothetical prote
115	6	3.7	187	2	S20526	hypothetical prote
116	6	3.7	188	1	JC2393	phospholipase A2 i
117	6	3.7	194	2	T51404	ribosomal protein-
118	6	3.7	195	2	T32997	hypothetical prote
119	6	3.7	199	2	S26355	opsin, 556nm - whi
120	6	3.7	199	2	S26356	opsin, 543nm - whi
121	6	3.7	201	1	A69999	phenylalanyl-tRNA
122	6	3.7	202	2	AG0124	probable prepilin
123	6	3.7	205	2	F81315	probable membrane
124	6	3.7	209	2	A84991	50S ribosomal prot
125	6	3.7	213	2	AD0696	riboflavin synthas
126	6	3.7	213	2	AC3389	uracil-DNA glycosy
127	6	3.7	215	1	A34588	DNA-directed RNA p
128	6	3.7	216	2	AF0789	AlkB protein [impo
129	6	3.7	216	2	T17693	hypothetical prote
130	6	3.7	216	2	AH2317	hypothetical prote
131	6	3.7	217	1	A24651	chloramphenicol O-
132	6	3.7	217	2	D70483	hypothetical prote
133	6	3.7	217	2	S50926	VTI1 protein - yea
134	6	3.7	218	2	A64124	GTP cyclohydrolase
135	6	3.7	219	1	QRECGP	glutamine transpor
136	6	3.7	219	2	H90739	glutamine transpor
137	6	3.7	219	2	B85590	glutamine transpor
138	6	3.7	219	2	AB0601	glutamine transpor
139	6	3.7	219	2	C84717	hypothetical prote
140	6	3.7	219	2	G64009	hypothetical prote
141	6	3.7	221	2	H86860	protein serine/thr
142	6	3.7	221	2	A70831	hypothetical prote
143	6	3.7	222	2	T30423	hypothetical prote
144	6	3.7	225	1	MNNZVT	nonstructural prot
145	6	3.7	225	1	MNNZP2	nonstructural prot
146	6	3.7	226	2	B86809	hypothetical prote
147	6	3.7	227	2	AC0482	hypothetical prote
148	6	3.7	227	2	T47186	hypothetical prote
149	6	3.7	231	1	TRPGTR	trypsin (EC 3.4.21
150	6	3.7	232	2	E75547	hypothetical prote
151	6	3.7	232	2	T16873	hypothetical prote
152	6	3.7	235	2	D75166	oxidoreductase PAB
153	6	3.7	235	2	S03839	superoxide dismuta
154	6	3.7	238	2	C35243	collagen alpha 2(V
155	6	3.7	238	2	E87055	probable conserved
156	6	3.7	239	2	D70854	hypothetical prote
157	6	3.7	241	2	S07899	probable fimbrial
158	6	3.7	242	2	A69026	conserved hypothet
159	6	3.7	243	2	S13296	Schwannoma-derived
160	6	3.7	243	2	D97189	glycosyltransferas
161	6	3.7	247	2	T39104	hypothetical prote
162	6	3.7	248	2	B90890	hypothetical prote
163	6	3.7	248	2	F91197	EspF protein [impo
164	6	3.7	248	2	B86044	espF [imported] -
165	6	3.7	250	1	MNVUPT	nonstructural prot
166	6	3.7	250	2	T33879	hypothetical prote
167	6	3.7	251	2	S72791	hypothetical prote
168	6	3.7	252	2	A87535	probable short cha
169	6	3.7	252	2	H83062	molybdopterin bios
170	6	3.7	254	2	S76459	transposase sll186
171	6	3.7	255	2	AF0515	DedA family integr
172	6	3.7	255	2	AE3551	tetratricopeptide
173	6	3.7	256	2	C70394	ABC transporter (A
174	6	3.7	256	2	D83336	probable thiol-dis
175	6	3.7	257	2	AE1533	oxidoreductase hom
176	6	3.7	258	2	F83207	probable permease
177	6	3.7	259	2	S76643	transposase slr051
178	6	3.7	260	2	T40430	conserved hypothet
179	6	3.7	261	2	S77171	transposase sll171
180	6	3.7	261	2	S77351	transposase sll171
181	6	3.7	261	2	S75081	transposase slr026
182	6	3.7	261	2	S76309	transposase slr035
183	6	3.7	261	2	C64959	probable export pr
184	6	3.7	261	2	A99965	flagellar biosynth
185	6	3.7	261	2	A85813	flagellar biosynth
186	6	3.7	262	2	B86662	hypothetical prote
187	6	3.7	262	2	T33597	hypothetical prote
188	6	3.7	263	2	S74989	hypothetical prote
189	6	3.7	264	2	A82253	tRNA pseudouridine
190	6	3.7	265	2	H69734	PBSX terminase sma
191	6	3.7	266	2	D82581	probable signal pe
192	6	3.7	270	2	T06118	hypothetical prote
193	6	3.7	275	2	H89880	hypothetical prote
194	6	3.7	275	2	T04675	hypothetical prote
195	6	3.7	275	2	T22040	hypothetical prote
196	6	3.7	276	2	G64097	fumarate reductase
197	6	3.7	276	2	T15536	hypothetical prote
198	6	3.7	277	2	I40410	PBSX prophage prot
199	6	3.7	277	2	B69269	conserved hypothet
200	6	3.7	278	2	D85171	OBP33pep like prot
201	6	3.7	283	2	S38230	hypothetical 32.9K
202	6	3.7	286	2	AF0610	probable pirin-rel
203	6	3.7	290	4	S07060	opsin, red-sensiti
204	6	3.7	291	2	D96002	probable sugar upt
205	6	3.7	291	2	B70928	probable dacB prot
206	6	3.7	292	2	E97004	uncharacterized co
207	6	3.7	293	2	T27430	hypothetical prote
208	6	3.7	294	2	T10445	peroxidase (EC 1.1
209	6	3.7	296	2	C95326	AttA2-like ABC tra
210	6	3.7	296	2	I40328	serum-resistance p
211	6	3.7	297	2	C84282	hypothetical prote
212	6	3.7	297	2	B86161	F1003.14 protein -
213	6	3.7	298	2	F90495	conserved hypothet
214	6	3.7	298	2	S72880	hypothetical prote
215	6	3.7	298	2	AB3056	O-linked GlcNAc tr
216	6	3.7	299	2	T32392	hypothetical prote
217	6	3.7	299	2	T16252	hypothetical prote
218	6	3.7	301	1	QQEC3R	transcription acti
219	6	3.7	301	2	B90631	transcription acti
220	6	3.7	301	2	B85482	transcription acti
221	6	3.7	301	2	H70505	hypothetical prote
222	6	3.7	301	2	A70731	probable glycerolp
223	6	3.7	303	2	AE3357	indole-3-glycerol
224	6	3.7	303	2	A97185	rhamnosyl transfer
225	6	3.7	305	2	B69548	molybdopterin oxid
226	6	3.7	305	2	B84779	hypothetical prote
227	6	3.7	306	2	E72598	probable ABC-trans
228	6	3.7	308	2	A84566	Ac-like transposas
229	6	3.7	309	2	F72316	hypothetical prote
230	6	3.7	309	2	AI3026	hypothetical prote
231	6	3.7	310	2	A83353	probable binding-p
232	6	3.7	312	2	E38245	transcription fact
233	6	3.7	313	2	AD0440	probable araC-fami
234	6	3.7	314	2	F64916	hypothetical prote
235	6	3.7	314	2	F90917	hypothetical prote
236	6	3.7	314	2	C85766	hypothetical prote
237	6	3.7	314	2	AC0683	probable exported
238	6	3.7	317	2	AB0281	probable exported
239	6	3.7	318	2	C84212	immunogenic protei
240	6	3.7	318	2	E82493	conserved hypothet
241	6	3.7	318	2	C38245	transcription fact
242	6	3.7	318	2	H98257	bifunctional carbo
243	6	3.7	319	2	F97626	probable transcrip
244	6	3.7	319	2	AH2849	transcription regu
245	6	3.7	324	2	S58142	coat protein - pha
246	6	3.7	325	2	D87669	Sua5/YciO/YrdC fam
247	6	3.7	326	2	G85727	probable transport
248	6	3.7	326	2	T50128	hypothetical serin

249	6	3.7	330	2	C75192
250	6	3.7	333	2	AH2038
251	6	3.7	335	2	H82046
252	6	3.7	336	2	H81091
253	6	3.7	336	2	C98230
254	6	3.7	338	1	R5HS3L
255	6	3.7	338	2	D69405
256	6	3.7	338	2	T23351
257	6	3.7	340	2	S76294
258	6	3.7	340	2	A64902
259	6	3.7	340	2	D38245
260	6	3.7	341	2	A35777
261	6	3.7	344	1	KHPGD
262	6	3.7	344	2	B90264
263	6	3.7	345	2	AH0457
264	6	3.7	345	2	E86202
265	6	3.7	346	2	H70722
266	6	3.7	346	2	A38245
267	6	3.7	347	2	AE3306
268	6	3.7	347	2	A39123
269	6	3.7	349	2	A71212
270	6	3.7	350	2	D71273
271	6	3.7	350	2	S43497
272	6	3.7	350	2	D72259
273	6	3.7	351	2	H90178
274	6	3.7	351	2	H87674
275	6	3.7	352	1	S58293
276	6	3.7	353	2	AI1859
277	6	3.7	355	2	D69049
278	6	3.7	356	2	F70397
279	6	3.7	356	2	A86213
280	6	3.7	356	2	A98250
281	6	3.7	356	2	AB3036
282	6	3.7	359	2	S40746
283	6	3.7	364	1	OOHUR
284	6	3.7	364	1	OOHUG
285	6	3.7	365	2	AB2902
286	6	3.7	366	2	D97677
287	6	3.7	367	2	T10207
288	6	3.7	367	2	D72376
289	6	3.7	370	2	S69718
290	6	3.7	371	2	S32604
291	6	3.7	371	2	C90176
292	6	3.7	375	2	C71917
293	6	3.7	375	2	H64597
294	6	3.7	375	2	B70500
295	6	3.7	376	2	F70576
296	6	3.7	376	2	S52137
297	6	3.7	379	2	T04469
298	6	3.7	382	2	H90127
299	6	3.7	383	1	G69201
300	6	3.7	384	2	B30319
301	6	3.7	386	2	G86652
302	6	3.7	388	2	S49445
303	6	3.7	388	2	E97260
304	6	3.7	391	2	I49141
305	6	3.7	391	2	A30319
306	6	3.7	391	2	S21335
307	6	3.7	391	2	JN0555
308	6	3.7	391	2	A38611
309	6	3.7	392	1	B69852
310	6	3.7	392	2	B69411
311	6	3.7	394	2	F86190
312	6	3.7	394	2	T10200
313	6	3.7	394	2	T20633
314	6	3.7	395	1	RRNZVT
315	6	3.7	395	1	RRNZP2
316	6	3.7	395	2	AB2977
317	6	3.7	395	2	C97244
318	6	3.7	396	2	AI3554
319	6	3.7	397	1	KIZYG
320	6	3.7	397	2	D75387
321	6	3.7	405	2	E81946

probable 1-aminocy
hypothetical prote
transcription repr
probable CDP-6-deo
hypothetical prote
ribosomal protein
Na+/H+ antiporter
hypothetical prote
hypothetical prote
peptide transport
transcription fact
transcription fact
cathepsin D (EC 3.
hypothetical prote
UDP-N-acetylmurama
hypothetical prote
probable transfera
transcription fact
dipeptide transpor
transcription fact
probable oligopept
probable glycerald
opsin, 563nm - whi
hypothetical prote
phosphate regulato
peptidase, M23/M37
myb-related protei
leucine dehydrogen
conserved hypothet
glutamate 5-kinase
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
opsin, red-sensiti
opsin, green-sensi
O-sialoglycoprotei
probable o-sialogl
oleoyl-[acyl-carri
pectate lyase - Th
hypothetical prote
collagen alpha 2(V
thiamin biosynthes
probable transamin
probable transamin
hypothetical prote
hypothetical prote
MID2 protein - yea
hypothetical prote
Mrs2p [imported] -
conserved hypothet
casein kinase II (
GTP-binding protei
RING finger protei
uncharacterized co
casein kinase (EC
casein kinase II (
casein kinase II (
casein kinase II (
casein kinase II (
probable NADH2 deh
3-ketoacyl-CoA thi
hypothetical prote
hypothetical prote
hypothetical prote
polymerase-associa
polymerase-associa
hippurate hydrolas
NADH:flavin oxidor
xylose transport s
phosphoglycerate k
NADH2 dehydrogenas
hypothetical prote

322	6	3.7	405	2	B97171
323	6	3.7	406	2	T19887
324	6	3.7	407	2	T48308
325	6	3.7	407	2	H81160
326	6	3.7	408	2	A71657
327	6	3.7	408	2	B75635
328	6	3.7	411	2	T34585
329	6	3.7	412	2	D97795
330	6	3.7	412	2	G64059
331	6	3.7	412	2	H81415
332	6	3.7	413	2	S04655
333	6	3.7	414	2	T16602
334	6	3.7	415	2	G72335
335	6	3.7	415	2	T32490
336	6	3.7	416	2	B84710
337	6	3.7	417	2	D72372
338	6	3.7	418	2	S74875
339	6	3.7	422	2	A98306
340	6	3.7	423	2	AB0054
341	6	3.7	425	2	A69770
342	6	3.7	425	2	A90055
343	6	3.7	426	2	E70601
344	6	3.7	426	2	F72402
345	6	3.7	428	2	AF0302
346	6	3.7	428	2	S48836
347	6	3.7	428	2	T05253
348	6	3.7	428	2	C87286
349	6	3.7	433	2	T14329
350	6	3.7	434	2	A71559
351	6	3.7	434	2	H81709
352	6	3.7	435	2	F86730
353	6	3.7	437	2	S18407
354	6	3.7	437	2	JC4988
355	6	3.7	437	2	C35147
356	6	3.7	437	2	T47831
357	6	3.7	438	2	A98161
358	6	3.7	438	2	H86006
359	6	3.7	438	2	AD0481
360	6	3.7	438	2	AB0997
361	6	3.7	445	2	S64418
362	6	3.7	446	2	E85691
363	6	3.7	448	2	AF1942
364	6	3.7	449	2	E87639
365	6	3.7	450	2	C70515
366	6	3.7	450	2	AD3117
367	6	3.7	452	2	G70905
368	6	3.7	452	2	F70552
369	6	3.7	453	2	A98170
370	6	3.7	454	2	AB2512
371	6	3.7	455	1	T21089
372	6	3.7	455	1	A29939
373	6	3.7	455	2	S05573
374	6	3.7	456	2	D70772
375	6	3.7	457	2	F83064
376	6	3.7	462	1	S36123
377	6	3.7	462	2	C85950
378	6	3.7	462	2	H91104
379	6	3.7	467	1	PWBSBF
380	6	3.7	467	2	C84420
381	6	3.7	468	2	T26081
382	6	3.7	469	2	S17726
383	6	3.7	470	2	D84863
384	6	3.7	473	2	B82765
385	6	3.7	476	2	T33273
386	6	3.7	481	2	AH0410
387	6	3.7	482	1	EXRT
388	6	3.7	482	2	T43996
389	6	3.7	483	2	A55033
390	6	3.7	484	2	T44183
391	6	3.7	484	2	T36538
392	6	3.7	484	2	JC7350
393	6	3.7	485	2	D91182
394	6	3.7	485	2	H86028

uncharacterized co
hypothetical prote
60S ribosomal prot
hypothetical prote
dihydrolipoamide a
phosphoenolpyruvat
probable secreted
hypothetical prote
probable serine tr
probable molybdenu
transcription fact
hypothetical prote
hypothetical prote
hypothetical prote
probable C3HC4 zin
hypothetical prote
quinolone resistan
probable hydrolase
probable Na+ depen
conserved hypothet
hypothetical prote
probable MoeAprote
UDP-N-acetylmuramo
probable solute-bi
cytochrome-c3 hydr
CER2 protein homol
Xaa-Pro dipeptidas
dermal glycoprotei
probable transcrip
N utilization subs
glutathione-disulf
acrosin (EC 3.4.21
high-affinity gluc
integrase homolog
hypothetical prote
hypothetical prote
hypothetical prote
probable gluconate
high-affinity gluc
hypothetical prote
probable tail comp
hypothetical prote
hypothetical prote
probable glnA3 pro
polygalacturonase
hypothetical prote
probable lpqF prot
hypothetical prote
replicative DNA he
acid phosphatase (
epoxide hydrolase
deoxyribodipyrimid
hypothetical prote
probable MFS trans
phosphotransferase
hypothetical prote
phosphotransferase
H+-transporting tw
hypothetical prote
hypothetical prote
H+-transporting tw
hypothetical prote
conserved hypothet
hypothetical prote
global stress requ
coagulation factor
virion protein [im
keratin 12 - mouse
probable envelope
probable transmemb
N-acetylglucosamin
probable outer mem
probable fimbrial

395	6	3.7	485	2	AE1221
396	6	3.7	485	2	AH1574
397	6	3.7	487	2	T05117
398	6	3.7	487	2	T00779
399	6	3.7	491	2	A70681
400	6	3.7	492	2	T38093
401	6	3.7	493	2	B83071
402	6	3.7	495	1	Q1AD52
403	6	3.7	495	2	T12401
404	6	3.7	496	1	Q1AD55
405	6	3.7	496	2	H70839
406	6	3.7	499	2	C71563
407	6	3.7	499	2	B75577
408	6	3.7	500	2	G71255
409	6	3.7	500	2	T26786
410	6	3.7	501	2	B75462
411	6	3.7	504	2	T05543
412	6	3.7	508	2	S74848
413	6	3.7	509	2	A96749
414	6	3.7	510	2	H81369
415	6	3.7	512	2	T09330
416	6	3.7	512	2	AF3402
417	6	3.7	513	2	A87324
418	6	3.7	514	2	H70699
419	6	3.7	522	2	T45540
420	6	3.7	523	2	C70717
421	6	3.7	524	2	T07988
422	6	3.7	525	2	G84406
423	6	3.7	525	2	E83913
424	6	3.7	527	2	T45439
425	6	3.7	530	2	T16972
426	6	3.7	531	2	T20763
427	6	3.7	533	2	H86749
428	6	3.7	539	2	S67049
429	6	3.7	541	1	WZEPPY
430	6	3.7	545	2	D90159
431	6	3.7	548	2	G82286
432	6	3.7	555	2	T24671
433	6	3.7	558	2	F64235
434	6	3.7	563	2	A88515
435	6	3.7	568	1	WZWCCC
436	6	3.7	568	2	S28792
437	6	3.7	572	2	AC0486
438	6	3.7	573	2	T25397
439	6	3.7	574	2	A84782
440	6	3.7	575	2	B83586
441	6	3.7	576	2	A48765
442	6	3.7	577	2	C84136
443	6	3.7	578	2	G82950
444	6	3.7	579	2	B44857
445	6	3.7	580	2	F84828
446	6	3.7	580	2	F84471
447	6	3.7	582	1	HNNZMP
448	6	3.7	582	1	HNNZMM
449	6	3.7	582	1	HNNZSB
450	6	3.7	582	2	S38184
451	6	3.7	582	2	A32034
452	6	3.7	583	2	A32861
453	6	3.7	587	2	C98165
454	6	3.7	587	2	AD3122
455	6	3.7	590	1	A54372
456	6	3.7	592	2	E86754
457	6	3.7	592	2	B81009
458	6	3.7	592	2	H82031
459	6	3.7	593	2	S45281
460	6	3.7	594	2	T31744
461	6	3.7	596	2	T23685
462	6	3.7	600	2	F75424
463	6	3.7	605	2	F35216
464	6	3.7	606	2	T29190
465	6	3.7	607	2	T20796
466	6	3.7	607	2	S77092
467	6	3.7	610	2	T22909

two-component sens
two-component sens
hypothetical prote
hypothetical prote
probable dctA prot
probable protein d
hypothetical prote
early E1B 55K prot
NADH2 dehydrogenas
early E1B 55K prot
hypothetical glyci
probable leucyl am
beta lactamase-rel
probable sodium/pr
hypothetical prote
hypothetical prote
hypothetical prote
neopullulanase - S
unknown protein T1
purH bifunctional
XIRF1 protein - hu
exopolyphosphatase
hypothetical prote
probable ppp prote
moaA protein [impo
probable purH prot
ATP-dependent clp
halolysin [importe
hypothetical prote
probable phosphori
probable high affi
hypothetical prote
hypothetical prote
probable membrane
pectate lyase (EC
hypothetical prote
phosphate ABC tran
hypothetical prote
Na+ ATPase chain J
polypeptide N-acet
pectate lyase (EC
pectate lyase (EC
pectate lyase (EC
hypothetical prote
hypothetical prote
probable permease
G protein-coupled
ABC transporter (A
conserved hypothet
acetolactate synth
probable laccase (
hypothetical prote
hemagglutinin-neur
hemagglutinin-neur
hemagglutinin-neur
hypothetical prote
choline kinase (EC
type II site-speci
probable ABC trans
hypothetical prote
G protein-coupled
prophage pi2 prote
BirA protein/Bvg a
probable biotin-[a
coagulation factor
hypothetical prote
hypothetical prote
probable cell cycl
FPD6 protein - fow
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote

468	6	3.7	610	2	F82192
469	6	3.7	610	2	S05807
470	6	3.7	612	2	T42243
471	6	3.7	614	2	T20795
472	6	3.7	615	1	KFHU12
473	6	3.7	619	2	I64087
474	6	3.7	623	2	T52656
475	6	3.7	623	2	A33622
476	6	3.7	623	2	T28958
477	6	3.7	636	2	T23232
478	6	3.7	637	1	QQVZ11
479	6	3.7	637	2	F72162
480	6	3.7	637	2	E42515
481	6	3.7	637	2	C36847
482	6	3.7	637	2	T28534
483	6	3.7	637	2	T37379
484	6	3.7	637	2	JH0611
485	6	3.7	639	2	JT0479
486	6	3.7	639	2	JQ0607
487	6	3.7	640	1	ALASGR
488	6	3.7	640	2	A29166
489	6	3.7	643	2	E82481
490	6	3.7	648	2	AB3013
491	6	3.7	650	2	JC4673
492	6	3.7	653	2	D69815
493	6	3.7	656	2	AE1479
494	6	3.7	662	2	E98271
495	6	3.7	664	1	TNBE70
496	6	3.7	667	2	S64915
497	6	3.7	670	2	T02019
498	6	3.7	671	2	S51599
499	6	3.7	672	2	I40333
500	6	3.7	674	2	D81220
501	6	3.7	674	2	F81990
502	6	3.7	681	2	T39814
503	6	3.7	685	2	B69401
504	6	3.7	691	2	A54741
505	6	3.7	693	1	TNBEF7
506	6	3.7	698	2	A69222
507	6	3.7	702	2	T39483
508	6	3.7	703	1	WMBET7
509	6	3.7	710	1	C43497
510	6	3.7	710	1	I42517
511	6	3.7	710	1	H36848
512	6	3.7	710	2	E72164
513	6	3.7	710	2	T28549
514	6	3.7	710	2	T37394
515	6	3.7	718	2	A82910
516	6	3.7	719	2	A87292
517	6	3.7	720	2	T43327
518	6	3.7	721	2	A39707
519	6	3.7	722	2	B71728
520	6	3.7	727	2	A88131
521	6	3.7	727	2	T08920
522	6	3.7	732	1	S36089
523	6	3.7	734	2	B82959
524	6	3.7	737	2	T46040
525	6	3.7	744	2	A82822
526	6	3.7	746	2	F95890
527	6	3.7	754	2	F87185
528	6	3.7	754	2	D70536
529	6	3.7	759	2	T39090
530	6	3.7	761	2	B97163
531	6	3.7	763	2	A29562
532	6	3.7	763	2	A47563
533	6	3.7	766	2	T20003
534	6	3.7	766	2	T01817
535	6	3.7	769	2	F81742
536	6	3.7	770	2	G02228
537	6	3.7	783	2	T37457
538	6	3.7	788	2	H96545
539	6	3.7	789	2	G90587
540	6	3.7	808	2	T51138

ABC transporter, A
SAN1 protein - yea
probable polypepti
hypothetical prote
coagulation factor
translation elonga
phosphoglucomutase
SAC1 protein - yea
hypothetical prote
hypothetical prote
early transcriptio
F6R protein - vari
D6R protein - vacc
early transcriptio
hypothetical prote
VETF-1, early tran
glutamate-cysteine
glucan 1,4-alpha-g
glucan 1,4-alpha-g
glucan 1,4-alpha-g
glucan 1,4-alpha-g
methyl-accepting c
metalloproteinase
protein kinase (EC
conserved hypothet
probable cell surf
metalloproteinase
70.5K alpha trans-
EMP70 protein prec
callus-associated
Om(2D) protein - f
tracheal colonizat
NADH dehydrogenase
NADH2 dehydrogenas
hypothetical prote
conserved hypothet
erythrocyte membra
73.8K alpha trans-
hypothetical prote
probable transmemb
UL17 protein - hum
early transcriptio
early transcriptio
early transcriptio
A8L protein - vari
hypothetical prote
VETF, 82K subunit
ppGpp 3'-pyrophosp
primosomal protein
gluconate transpor
erythrocyte membra
probable peptidase
protein F10G7.9 [i
hypothetical prote
translation elonga
hypothetical prote
hypothetical prote
NADH-ubiquinone ox
probable aldehyde
phosphoribosylform
probable purL prot
probable integral
ATP-dependent seri
PRT1 protein - yea
glucose-6-phosphat
hypothetical prote
hypothetical prote
conserved hypothet
DOC-2 - human
Tricorn proteinase
hypothetical prote
lipoprotein [impor
probable glutamate

541	6	3.7	820	2	H86246	hypothetical prote
542	6	3.7	825	2	T23612	hypothetical prote
543	6	3.7	834	2	S54563	hypothetical prote
544	6	3.7	835	2	F90260	hypothetical prote
545	6	3.7	839	2	A56337	glycoprotein phosp
546	6	3.7	840	2	T02164	hypothetical prote
547	6	3.7	847	2	C87678	TonB-dependent rec
548	6	3.7	849	2	C90834	tail length tape m
549	6	3.7	849	2	T22306	hypothetical prote
550	6	3.7	850	2	S54553	SIN3 protein-bindi
551	6	3.7	853	1	TLBPHL	minor tail protein
552	6	3.7	869	2	S53098	envelope polyprote
553	6	3.7	878	2	A83748	endo-beta-N-acetyl
554	6	3.7	879	2	T02728	serine/threonine-s
555	6	3.7	905	2	F82734	serine proteinase
556	6	3.7	910	2	H82826	serine proteinase
557	6	3.7	917	2	S09646	collagen alpha 2(V
558	6	3.7	920	2	I40614	surface array prot
559	6	3.7	921	2	A48184	transcription init
560	6	3.7	921	2	A45183	TBP-associated fac
561	6	3.7	925	2	T06128	hypothetical prote
562	6	3.7	930	2	D86546	polymorphic outer
563	6	3.7	930	2	D72078	polymorphic outer
564	6	3.7	930	2	A81591	polymorphic membra
565	6	3.7	931	2	T43725	ubiquitin fusion d
566	6	3.7	934	2	S75633	hypothetical prote
567	6	3.7	936	2	E71405	probable ankyrin -
568	6	3.7	942	2	D96814	trehalose-6-phosph
569	6	3.7	948	2	A57640	retinoblastoma bin
570	6	3.7	951	2	T00017	gene ADAMTS-1 prot
571	6	3.7	967	2	F97442	isoleucyl-tRNA syn
572	6	3.7	967	2	AH2660	isoleucyl-tRNA syn
573	6	3.7	969	2	S37886	hypothetical prote
574	6	3.7	978	2	T16948	hypothetical prote
575	6	3.7	989	2	AE2140	toxin secretion AB
576	6	3.7	997	2	S33754	glutamate receptor
577	6	3.7	1003	2	H82883	hypothetical prote
578	6	3.7	1010	2	T09499	ATP-dependent clp
579	6	3.7	1014	2	S75724	hypothetical prote
580	6	3.7	1017	2	T30195	DNA-directed DNA p
581	6	3.7	1017	2	F82500	vgrG protein VCA01
582	6	3.7	1018	1	CGHU2A	collagen alpha 2(V
583	6	3.7	1026	2	T18220	chitin synthase (E
584	6	3.7	1029	1	S21369	collagen alpha 2(V
585	6	3.7	1029	2	T38951	probable ubiquitin
586	6	3.7	1034	2	S63536	aminomethyltransfe
587	6	3.7	1035	2	T07826	aminomethyltransfe
588	6	3.7	1037	2	A36096	Ca2+-transporting
589	6	3.7	1053	2	S46199	probable complemen
590	6	3.7	1057	2	T04874	hypothetical prote
591	6	3.7	1057	2	A42109	glycine dehydrogen
592	6	3.7	1060	2	S33641	homeotic protein z
593	6	3.7	1081	2	S15040	pleiotropic drug r
594	6	3.7	1086	2	AH2136	microcystin synthe
595	6	3.7	1097	1	S47220	protein kinase C (
596	6	3.7	1102	2	S35617	probable origin-bi
597	6	3.7	1121	2	C82120	transcription regu
598	6	3.7	1124	2	T30340	dsRNA adenosine de
599	6	3.7	1127	2	T25804	hypothetical prote
600	6	3.7	1131	2	F88570	protein K03H1.2 [i
601	6	3.7	1134	2	T04587	hypothetical prote
602	6	3.7	1139	2	S40932	hypothetical prote
603	6	3.7	1166	2	T28680	fibrinogen-binding
604	6	3.7	1174	2	A40853	potassium channel
605	6	3.7	1185	2	T19212	hypothetical prote
606	6	3.7	1203	2	T04294	hypothetical prote
607	6	3.7	1203	2	T21275	hypothetical prote
608	6	3.7	1221	2	A49457	fibulin-2 precurso
609	6	3.7	1240	2	T03097	CDO protein - huma
610	6	3.7	1251	2	T21389	hypothetical prote
611	6	3.7	1270	2	T51227	related to verruco
612	6	3.7	1290	2	A56493	leucocyte common a
613	6	3.7	1315	2	A56101	collagen alpha 1(X
614	6	3.7	1363	2	C84346	hypothetical prote
615	6	3.7	1379	2	JC5778	apoptosis signal-r
616	6	3.7	1387	2	A96771	hypothetical prote
617	6	3.7	1390	2	T31353	polyprotein - Arab
618	6	3.7	1391	2	S73652	RNA polymerase bet
619	6	3.7	1398	2	C87448	hypothetical prote
620	6	3.7	1400	1	I38185	protein-tyrosine k
621	6	3.7	1405	2	H87230	probable integral
622	6	3.7	1407	1	T00558	probable ABC trans
623	6	3.7	1408	2	T47671	P-glycoprotein-lik
624	6	3.7	1417	2	A83080	hypothetical prote
625	6	3.7	1440	2	T44872	probable integral
626	6	3.7	1451	2	S41025	hypothetical prote
627	6	3.7	1458	2	S36014	dynein heavy chain
628	6	3.7	1473	2	T13855	suppressor of sabl
629	6	3.7	1492	2	T18560	DNA-directed DNA p
630	6	3.7	1494	2	T13798	hypothetical prote
631	6	3.7	1500	2	AF2027	hypothetical prote
632	6	3.7	1508	2	T27828	hypothetical prote
633	6	3.7	1513	2	T28158	probable DNA-direc
634	6	3.7	1519	2	T27829	hypothetical prote
635	6	3.7	1560	2	T42727	proliferation pote
636	6	3.7	1633	2	T01879	hypothetical prote
637	6	3.7	1653	2	B91052	hypothetical prote
638	6	3.7	1653	2	F85896	hypothetical prote
639	6	3.7	1653	2	G65028	hypothetical prote
640	6	3.7	1691	1	A44212	genome polyprotein
641	6	3.7	1708	1	A43100	ataxia telangiecta
642	6	3.7	1761	2	T13990	collagen type IV a
643	6	3.7	1771	2	S76851	hypothetical prote
644	6	3.7	1774	2	B56101	collagen alpha 1(X
645	6	3.7	1857	2	T50513	hypothetical prote
646	6	3.7	1897	1	TDHULK	leukocyte antigen-
647	6	3.7	1984	2	T13171	probable vitelloge
648	6	3.7	2136	2	A05037	hypothetical prote
649	6	3.7	2153	2	T14893	scavenger receptor
650	6	3.7	2273	2	T09083	hemagglutinin/hemo
651	6	3.7	2340	2	B71704	cell surface antig
652	6	3.7	2514	1	MNWVN2	nonstructural poly
653	6	3.7	2783	2	T31431	inositol 1,4,5-tri
654	6	3.7	2870	2	H96974	cyclic beta 1-2 gl
655	6	3.7	2958	2	S64921	probable membrane
656	6	3.7	3013	2	AB0480	probable invasin Y
657	6	3.7	3029	2	S76109	hypothetical prote
658	6	3.7	3149	1	QQBE8	BPLF1 protein - hu
659	6	3.7	3449	2	T01083	hypothetical prote
660	6	3.7	3600	2	D86161	F1003.12 protein -
661	6	3.7	4085	2	S28600	hypothetical prote
662	6	3.7	4351	2	T00252	MEGF1 protein - ra
663	6	3.7	4644	1	A38905	dynein heavy chain
664	6	3.7	4725	1	A44357	dynein heavy chain
665	6	3.7	4936	2	AH2515	hypothetical prote
666	6	3.7	5005	2	F82884	hypothetical prote
667	6	3.7	5232	2	A45086	HC-toxin synthetas
668	5	3.1	14	2	B56884	Pax-QNR, long form
669	5	3.1	14	2	PT0223	Ig heavy chain CDR
670	5	3.1	15	2	PQ0778	NADH2 dehydrogenas
671	5	3.1	18	2	A28027	protein P2 - curle
672	5	3.1	24	2	A41037	antimicrobial pept
673	5	3.1	25	2	A34179	neuromedin U - com
674	5	3.1	27	2	C69350	hypothetical prote
675	5	3.1	29	2	A49410	t-complex polypept
676	5	3.1	29	2	B85840	hypothetical prote
677	5	3.1	29	2	C85840	hypothetical prote
678	5	3.1	30	2	A18780	dimethylallyltrans
679	5	3.1	32	2	F90807	hypothetical prote
680	5	3.1	32	2	B85667	hypothetical prote
681	5	3.1	32	2	A64848	hypothetical prote
682	5	3.1	35	2	E38601	Ig kappa chain V r
683	5	3.1	37	2	A81552	hypothetical prote
684	5	3.1	38	2	S78357	photosystem II pro
685	5	3.1	38	2	I65992	gene MLL protein -
686	5	3.1	43	2	C72270	hypothetical prote

687	5	3.1	44	2	B89984
688	5	3.1	45	2	I39208
689	5	3.1	46	2	S20018
690	5	3.1	47	2	G87679
691	5	3.1	48	2	S13571
692	5	3.1	49	2	S29173
693	5	3.1	49	2	F82535
694	5	3.1	50	2	AC2221
695	5	3.1	51	2	I40723
696	5	3.1	51	2	E82981
697	5	3.1	51	2	E48652
698	5	3.1	51	2	T36214
699	5	3.1	52	2	AI0472
700	5	3.1	52	2	AI1345
701	5	3.1	53	2	E95311
702	5	3.1	55	1	R5EC33
703	5	3.1	55	2	E82347
704	5	3.1	55	2	AD0007
705	5	3.1	55	2	AF0971
706	5	3.1	55	2	G91192
707	5	3.1	55	2	H86039
708	5	3.1	56	2	S41602
709	5	3.1	57	2	AD0065
710	5	3.1	58	2	C84363
711	5	3.1	58	2	E89468
712	5	3.1	60	2	B61510
713	5	3.1	61	2	D83531
714	5	3.1	61	2	T19658
715	5	3.1	61	2	C81531
716	5	3.1	63	2	G96531
717	5	3.1	63	2	E84370
718	5	3.1	63	2	T37351
719	5	3.1	63	2	T28506
720	5	3.1	63	2	B72159
721	5	3.1	64	2	C87511
722	5	3.1	64	2	B72740
723	5	3.1	64	2	AF3568
724	5	3.1	65	2	E84494
725	5	3.1	65	2	E97099
726	5	3.1	66	2	F90146
727	5	3.1	66	2	G98059
728	5	3.1	67	1	Q7BPE7
729	5	3.1	67	1	Q7BPT3
730	5	3.1	67	2	G82568
731	5	3.1	67	2	T42055
732	5	3.1	67	2	T07561
733	5	3.1	67	2	A71289
734	5	3.1	67	2	AH3199
735	5	3.1	68	2	T44551
736	5	3.1	68	2	AD2807
737	5	3.1	69	2	B83524
738	5	3.1	69	2	B83374
739	5	3.1	69	2	B69355
740	5	3.1	69	2	S41000
741	5	3.1	69	2	H82753
742	5	3.1	70	2	S35030
743	5	3.1	70	2	I39058
744	5	3.1	70	2	C97412
745	5	3.1	71	2	I51747
746	5	3.1	71	2	S74348
747	5	3.1	71	2	AF0683
748	5	3.1	71	2	AI2240
749	5	3.1	72	2	D87340
750	5	3.1	72	2	G83106
751	5	3.1	72	2	E83733
752	5	3.1	73	2	D86616
753	5	3.1	73	2	T44814
754	5	3.1	73	2	C72008
755	5	3.1	74	2	AD3290
756	5	3.1	76	2	C84094
757	5	3.1	76	2	T17960
758	5	3.1	77	2	S52636
759	5	3.1	77	2	T05712

hypothetical prote
H+-exporting ATPas
hypothetical prote
hypothetical prote
hypothetical prote
D-galactose-bindin
hypothetical prote
hypothetical prote
lysC protein - Cor
50S ribosomal prot
integral membrane
spdC protein - Str
hypothetical prote
hypothetical prote
hypothetical prote
ribosomal protein
ribosomal protein
50S ribosomal prot
50S ribosomal chai
50S ribosomal subu
50S ribosomal subu
interferon alpha r
hypothetical prote
ferredoxin [import
protein F07G6.5 [i
translation elonga
carbon storage reg
hypothetical prote
conserved hypothet
hypothetical prote
hypothetical prote
RNA polymerase sub
hypothetical prote
I6R protein - vari
hypothetical prote
probable ribonucle
hypothetical prote
hypothetical prote
hypothetical prote
DNA-directed RNA p
hypothetical prote
gene 17.5 protein
lysis protein - ph
cold shock protein
cold shock protein
hypothetical prote
probable DNA-direc
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
photosystem I chai
hypothetical prote
hypothetical prote
MHC class II alpha
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
CT849.1 hypothetic
hypothetical prote
CT849.1 hypothetic
hypothetical prote
protein translocat
hypothetical prote
metallothionein -
auxin-induced prot

760	5	3.1	77	2	H69328
761	5	3.1	77	2	C69902
762	5	3.1	77	2	A37317
763	5	3.1	78	2	A27683
764	5	3.1	79	1	TIEPVK
765	5	3.1	80	2	AB2803
766	5	3.1	80	2	AE3392
767	5	3.1	80	2	E84630
768	5	3.1	80	2	T17805
769	5	3.1	80	2	S19984
770	5	3.1	80	2	S19987
771	5	3.1	80	2	F98090
772	5	3.1	81	2	S41580
773	5	3.1	81	2	S22826
774	5	3.1	81	2	T35731
775	5	3.1	81	2	T47865
776	5	3.1	81	2	A24522
777	5	3.1	81	2	AB3462
778	5	3.1	81	2	AF2553
779	5	3.1	82	2	T52378
780	5	3.1	83	2	C95126
781	5	3.1	83	2	S70360
782	5	3.1	83	2	C69440
783	5	3.1	83	2	H85523
784	5	3.1	83	2	G97996
785	5	3.1	83	2	AD0081
786	5	3.1	83	2	AB2036
787	5	3.1	84	2	E86209
788	5	3.1	84	2	T17676
789	5	3.1	85	2	C39247
790	5	3.1	85	2	E30536
791	5	3.1	85	2	F30536
792	5	3.1	85	2	I68800
793	5	3.1	85	2	I68788
794	5	3.1	85	2	D39247
795	5	3.1	85	2	A39247
796	5	3.1	85	2	I68793
797	5	3.1	85	2	I68795
798	5	3.1	85	2	I68794
799	5	3.1	85	2	I68796
800	5	3.1	85	2	I68802
801	5	3.1	85	2	I68799
802	5	3.1	85	2	T03332
803	5	3.1	86	2	D33989
804	5	3.1	86	2	F34964
805	5	3.1	86	2	D64371
806	5	3.1	87	2	G30536
807	5	3.1	87	2	H30536
808	5	3.1	87	2	I59648
809	5	3.1	87	2	T34648
810	5	3.1	87	2	S15699
811	5	3.1	87	2	F40361
812	5	3.1	87	2	T43570
813	5	3.1	87	2	A71426
814	5	3.1	87	2	C97909
815	5	3.1	88	1	W4WL18
816	5	3.1	88	2	D90957
817	5	3.1	88	2	H86752
818	5	3.1	88	2	T06988
819	5	3.1	88	2	T41918
820	5	3.1	89	2	G83514
821	5	3.1	89	2	S72603
822	5	3.1	90	2	S29977
823	5	3.1	90	2	A87563
824	5	3.1	90	2	F34768
825	5	3.1	90	2	S45092
826	5	3.1	90	2	T15357
827	5	3.1	91	2	JQ0905
828	5	3.1	91	2	T09285
829	5	3.1	91	2	T09293
830	5	3.1	91	2	T17968
831	5	3.1	91	2	F70630
832	5	3.1	92	2	C90667

nifU protein (nifU
permease homolog y
probable membrane
heat shock 90K pro
K+ channel blocker
conserved hypothet
phosphoribosylform
probable small nuc
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
lysozyme (EC 3.2.1
small nuclear ribo
hypothetical prote
transport protein
mitochondrial prot
hypothetical prote
hypothetical prote
probable transport
hypothetical prote
exoD protein homol
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
protein F22G5.16 [
hypothetical prote
MHC class II histo
MHC class II histo
MHC class II histo
MHC class II histo
MHC class II histo
MHC class II histo
MHC class II histo
MHC class II histo
MHC class II histo
MHC class II histo
MHC class II histo
MHC class II histo
gene e35 protein -
Ig heavy chain V-4
Ig heavy chain V-I
hypothetical prote
MHC class II histo
MHC class II histo
HLA-DPB1 - human (
probable SIR2 fami
hypothetical prote
virC-region hypoth
type III secretion
hypothetical prote
hypothetical prote
E4 protein - human
hypothetical prote
prophage pi2 prote
pathogen-induced p
hypothetical prote
conserved hypothet
hypothetical prote
hypC protein - Alc
PsbAd protein [imp
ORF6 protein - Orf
probable copy cont
hypothetical prote
bombyxin B-2 homol
embryonic abundant
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical membr

833	5	3.1	92	2	A90764	hypothetical prote
834	5	3.1	92	2	A90819	hypothetical prote
835	5	3.1	92	2	S22829	copR protein - Str
836	5	3.1	92	2	G85517	hypothetical prote
837	5	3.1	93	1	WMIHB9	3c protein - avian
838	5	3.1	93	2	T10379	hypothetical prote
839	5	3.1	93	2	AE1247	hypothetical prote
840	5	3.1	93	2	AI1609	hypothetical prote
841	5	3.1	94	2	C96576	hypothetical prote
842	5	3.1	94	2	JH0207	hypothetical 10.8K
843	5	3.1	94	2	T50619	hypothetical prote
844	5	3.1	94	2	AG1365	hypothetical prote
845	5	3.1	95	2	S30068	polygalacturonase
846	5	3.1	95	2	S30069	polygalacturonase
847	5	3.1	95	2	S30062	polygalacturonase
848	5	3.1	95	2	G37262	Ig heavy chain V r
849	5	3.1	95	2	A41872	heat shock protein
850	5	3.1	95	2	G97232	co-chaperonin GroE
851	5	3.1	95	2	B84216	hypothetical prote
852	5	3.1	95	2	T45374	hypothetical prote
853	5	3.1	96	2	D47624	Ig heavy chain V-I
854	5	3.1	96	2	E86773	citrate (pro-3S)-l
855	5	3.1	96	2	G86801	prophage pi3 prote
856	5	3.1	96	2	F81890	hypothetical prote
857	5	3.1	96	2	S72376	pd96 protein - Ent
858	5	3.1	96	2	E69063	hypothetical prote
859	5	3.1	97	2	S75342	hypothetical prote
860	5	3.1	97	2	T45331	hypothetical prote
861	5	3.1	97	2	T38165	hypothetical prote
862	5	3.1	97	2	C97072	uncharacterized pr
863	5	3.1	98	2	A81133	hypothetical prote
864	5	3.1	98	2	JC5822	NADH2 dehydrogenas
865	5	3.1	99	2	S25971	hypothetical prote
866	5	3.1	99	2	T01678	hypothetical prote
867	5	3.1	100	1	BVMYBA	chaperonin groES -
868	5	3.1	100	1	BVMY7B	chaperonin groES -
869	5	3.1	100	2	A64076	urease (EC 3.5.1.5
870	5	3.1	100	2	S25180	heat shock protein
871	5	3.1	100	2	S62753	cerato-ulmin precu
872	5	3.1	100	2	A64016	hypothetical prote
873	5	3.1	100	2	C72375	hypothetical prote
874	5	3.1	101	1	PFHU4	platelet factor 4
875	5	3.1	101	2	T38824	thioltransferase -
876	5	3.1	101	2	D75013	hypothetical prote
877	5	3.1	101	2	H95941	hypothetical prote
878	5	3.1	101	2	D96576	hypothetical prote
879	5	3.1	101	2	S30975	gene 30 protein -
880	5	3.1	101	2	H75370	stage V sporulatio
881	5	3.1	102	1	JQ1300	calgizzarin - rabb
882	5	3.1	102	2	S37565	chaperonin groES -
883	5	3.1	102	2	A41325	heat shock protein
884	5	3.1	102	2	C87282	conserved hypothet
885	5	3.1	102	2	F69475	conserved hypothet
886	5	3.1	102	2	H82648	conserved hypothet
887	5	3.1	103	1	K4RB	Ig kappa-B4 chain
888	5	3.1	103	1	G64093	ribosomal protein
889	5	3.1	103	2	S04101	hemolysin A - Prot
890	5	3.1	103	2	B81160	conserved hypothet
891	5	3.1	103	2	S54839	phosphoribosyl-ATP
892	5	3.1	103	2	S64330	probable membrane
893	5	3.1	103	2	G42528	B26R protein - vac
894	5	3.1	103	2	B70600	hypothetical prote
895	5	3.1	104	1	PFHU4A	platelet factor 4
896	5	3.1	104	2	A71817	probable thioredox
897	5	3.1	104	2	B64702	thioredoxin - Heli
898	5	3.1	104	2	F53275	Ig kappa-1 chain C
899	5	3.1	104	2	B69058	hypothetical prote
900	5	3.1	105	1	I37080	calgizzarin - huma
901	5	3.1	105	2	C71635	ribosomal protein
902	5	3.1	105	2	B97845	50S ribosomal prot
903	5	3.1	105	2	A26774	platelet factor 4
904	5	3.1	105	2	S23061	chorion protein -
905	5	3.1	105	2	F70730	hypothetical prote
906	5	3.1	105	2	D85677	unknown protein en
907	5	3.1	105	2	H90817	hypothetical prote
908	5	3.1	105	2	D97507	hypothetical prote
909	5	3.1	105	2	AI3472	hypothetical prote
910	5	3.1	105	2	H71635	hypothetical prote
911	5	3.1	105	2	C97846	hypothetical prote
912	5	3.1	106	1	WMIH23	3c protein - avian
913	5	3.1	106	2	S28688	hypothetical prote
914	5	3.1	106	2	B71659	hypothetical prote
915	5	3.1	106	2	T24774	hypothetical prote
916	5	3.1	106	2	C70794	hypothetical prote
917	5	3.1	106	2	AB0256	probable phage pro
918	5	3.1	106	2	G69477	NADH2 dehydrogenas
919	5	3.1	107	1	WMIH26	3c protein - avian
920	5	3.1	107	2	S77775	probable CDPdiacyl
921	5	3.1	107	2	I68730	IgE chain C3 regio
922	5	3.1	107	2	I68726	IgE chain C3 regio
923	5	3.1	107	2	AG3490	hypothetical trans
924	5	3.1	107	2	E71634	glutaredoxin-like
925	5	3.1	107	2	G83788	hypothetical prote
926	5	3.1	107	2	B71069	hypothetical prote
927	5	3.1	107	2	S09879	hypothetical prote
928	5	3.1	108	1	H64913	probable membrane
929	5	3.1	108	2	H90914	hypothetical prote
930	5	3.1	108	2	C87494	conserved hypothet
931	5	3.1	108	2	H83992	hypothetical prote
932	5	3.1	108	2	E85763	hypothetical prote
933	5	3.1	108	2	F90033	conserved hypothet
934	5	3.1	108	2	AH0679	probable membrane
935	5	3.1	108	2	H86287	F9L1.29 protein -
936	5	3.1	108	2	AI2333	hypothetical prote
937	5	3.1	108	2	F64545	hypothetical prote
938	5	3.1	108	2	D72470	hypothetical prote
939	5	3.1	109	1	QQIHI1	3c protein - avian
940	5	3.1	109	2	B69806	conserved hypothet
941	5	3.1	109	2	AD2091	anti-sigma factor
942	5	3.1	110	1	R5BYA1	acidic ribosomal p
943	5	3.1	110	1	R6BY22	60s acidic ribosom
944	5	3.1	110	2	C70964	hypothetical prote
945	5	3.1	110	2	F71095	hypothetical prote
946	5	3.1	110	2	D64341	hypothetical prote
947	5	3.1	110	2	A72661	hypothetical prote
948	5	3.1	110	2	H64512	hypothetical prote
949	5	3.1	110	2	G70021	hypothetical prote
950	5	3.1	110	2	T50423	hypothetical prote
951	5	3.1	110	2	A38529	nikA protein - Esc
952	5	3.1	110	2	S75767	hypothetical prote
953	5	3.1	110	2	B69403	hypothetical prote
954	5	3.1	110	2	D82628	hypothetical prote
955	5	3.1	111	1	JX0120	ribonuclease-relat
956	5	3.1	111	2	B45403	insulin-like growt
957	5	3.1	111	2	I48640	neurotoxin homolog
958	5	3.1	111	2	C41839	ribosomal protein
959	5	3.1	111	2	G71288	conserved hypothet
960	5	3.1	111	2	D97843	glutaredoxin-like
961	5	3.1	111	2	G34768	hypothetical prote
962	5	3.1	111	2	S64364	probable membrane
963	5	3.1	111	2	B81106	hypothetical prote
964	5	3.1	112	2	S33180	nitrogen regulator
965	5	3.1	112	2	A86604	periplasmic divale
966	5	3.1	112	2	B72019	periplasmic divale
967	5	3.1	112	2	F70954	probable lsr2 prot
968	5	3.1	112	2	B43601	LSR2 T-cell antige
969	5	3.1	112	2	B48941	pediocin productio
970	5	3.1	113	2	S66512	ribosomal protein
971	5	3.1	113	2	S73951	hypothetical prote
972	5	3.1	113	2	F72687	hypothetical prote
973	5	3.1	113	2	AF3643	hypothetical prote
974	5	3.1	114	2	PQ0745	phytochrome - Meso
975	5	3.1	114	2	T30866	hypothetical prote
976	5	3.1	114	2	AF2315	hypothetical prote
977	5	3.1	114	2	S75565	hypothetical prote
978	5	3.1	114	2	H96541	hypothetical prote

979	5	3.1	115	2	B32071	T-cell receptor de
980	5	3.1	115	2	S16534	metallothionein-li
981	5	3.1	115	2	F84469	probable glycine-r
982	5	3.1	115	2	T20559	hypothetical prote
983	5	3.1	115	2	E83184	hypothetical prote
984	5	3.1	116	1	S34611	nitrogenase (EC 1.
985	5	3.1	116	2	S35468	NADH2 dehydrogenas
986	5	3.1	116	2	S39434	spermadhesin AQN-3
987	5	3.1	116	2	S17567	AQN-3 protein - pi
988	5	3.1	116	2	F64510	hypothetical prote
989	5	3.1	116	2	T15004	hypothetical prote
990	5	3.1	116	2	F82805	conserved hypothet
991	5	3.1	116	2	D75355	hypothetical prote
992	5	3.1	116	2	A98319	hypothetical prote
993	5	3.1	116	2	AD2964	hypothetical prote
994	5	3.1	117	1	VCWWRH	coat protein - rab
995	5	3.1	117	2	T41247	quinone oxidoreduc
996	5	3.1	117	2	T02872	probable lipid tra
997	5	3.1	117	2	A45387	developmental regu
998	5	3.1	117	2	C64116	ribosomal protein
999	5	3.1	117	2	B82476	ribosomal protein
1000	5	3.1	117	2	S22135	coat protein - rab

ALIGNMENTS

RESULT 1
S56545
fimbrial protein fimH precursor, type 1 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 01-Mar-2002
C;Accession: S56545; B65246; S09563; A36967
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A;Reference number: S56314; MUID:95334362; PMID:7610040
A;Accession: S56545
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-300 <BUR>
A;Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97216.1; PID:g537161
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B65246
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-300 <BLAT>
A;Cross-references: GB:AE000502; GB:U00096; NID:g2367374; PIDN:AAC77276.1; PID:g1790775;
A;Experimental source: strain K-12, substrain MG1655
R;Klemm, P.; Christiansen, G.
Mol. Gen. Genet. 208, 439-445, 1987
A;Title: Three fim genes required for the regulation of length and mediation of adhesion
A;Reference number: S07321; MUID:88038337; PMID:2890081
A;Accession: S09563
A;Molecule type: DNA
A;Residues: 1-196,'R',198-221,'H',223-300 <KLE>
A;Cross-references: EMBL:X05672; NID:g41463; PIDN:CAA29156.1; PID:g41466
R;Sokurenko, E.V.; Courtney, H.S.; Ohman, D.E.; Klemm, P.; Hasty, D.L.
J. Bacteriol. 176, 748-755, 1994
A;Title: FimH family of type 1 fimbrial adhesins: functional heterogeneity due to minor
A;Reference number: A36967; MUID:94131954; PMID:7905476
A;Accession: A36967
A;Status: not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-78,'R',80-221,'H',223-300 <SOK>
A;Note: sequence extracted from NCBI backbone (NCBIP:143314)
C;Genetics:
A;Gene: fimH
C;Function:
A;Description: involved in longitudinal regulation and mannose-specific adhesion
A;Note: not necessary for the production of fimbriae
A;Note: controls length and number of fimbriae
C;Superfamily: fimbrial protein fimH
C;Keywords: fimbria
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-300/Product: fimbrial protein fimH, type 1 #status predicted <MAT>

Query Match 57.1%; Score 92; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 5.6e-87;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db     107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138
```

RESULT 2
G91288
hypothetical protein ECs5279 [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: G91288
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G91288
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-300 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB38702.1; PID:g13364757; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs5279
C;Superfamily: fimbrial protein fimH

Query Match 55.9%; Score 90; DB 2; Length 300;
Best Local Similarity 100.0%; Pred. No. 6.5e-85;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||
Db     109 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 138
```

RESULT 3
B86130
hypothetical protein fimH [imported] - Escherichia coli (strain O157:H7, substrain ED
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B86130
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B86130
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-300 <STO>
A;Cross-references: GB:AE005174; NID:g12519327; PIDN:AAG59502.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933

```
C;Genetics:
A;Gene: fimH
C;Superfamily: fimbrial protein fimH

  Query Match             55.9%;  Score 90;  DB 2;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-85;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     49 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 108

Qy     63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||
Db    109 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 4
A32801
fimbrial adhesin precursor, type 1 - Klebsiella pneumoniae
C;Species: Klebsiella pneumoniae
C;Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 18-Sep-1998
C;Accession: A32801
R;Gerlach, G.F.; Clegg, S.; Allen, B.L.
J. Bacteriol. 171, 1262-1270, 1989
A;Title: Identification and characterization of the genes encoding the type 3 and type 1
A;Reference number: A32801; MUID:89155420; PMID:2563996
A;Accession: A32801
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-302 <GER>
A;Cross-references: GB:M24564
C;Superfamily: fimbrial protein fimH

  Query Match             26.1%;  Score 42;  DB 2;  Length 302;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-35;
  Matches   42;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLS 44
          |||||||||||||||||||||||||||||||||||||||||||
Db     51 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLS 92


RESULT 5
A64904
probable fimbrial protein b1502 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: A64904
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A64904
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-304 <BLAT>
A;Cross-references: GB:AE000247; GB:U00096; NID:g1787773; PIDN:AAC74575.1; PID:g1787779;
A;Experimental source: strain K-12, substrain MG1655
C;Superfamily: fimbrial protein fimH
C;Keywords: fimbria

  Query Match             5.6%;  Score 9;  DB 2;  Length 304;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 GQNLVVDLS 14
          |||||||||
Db     53 GQNLVVDLS 61


RESULT 6
C90892
probable adhesin [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050995
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: C90892
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90892
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-304 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35530.1; PID:g13361573; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2107
C;Superfamily: fimbrial protein fimH

  Query Match             5.6%;  Score 9;  DB 2;  Length 304;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 GQNLVVDLS 14
          |||||||||
Db     53 GQNLVVDLS 61


RESULT 7
F85725
probable adhesin, FimH type protein Z2206 [imported] - Escherichia coli (strain O157:
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: F85725
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85725
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-304 <STO>
A;Cross-references: GB:AE005174; NID:g12515169; PIDN:AAG56266.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z2206
C;Superfamily: fimbrial protein fimH

  Query Match             5.6%;  Score 9;  DB 2;  Length 304;
  Best Local Similarity   100.0%;  Pred. No. 0.32;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 GQNLVVDLS 14
          |||||||||
Db     53 GQNLVVDLS 61


RESULT 8
F75497
probable arginine utilization protein RocB - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75497
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75497
A;Status: preliminary
```

A;Molecule type: DNA
A;Residues: 1-539 <WHI>
A;Cross-references: GB:AE001919; GB:AE000513; NID:g6458307; PIDN:AAF10189.1; PID:g645830
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0612
A;Map position: 1
C;Superfamily: Deinococcus radiodurans probable arginine utilization protein RocB

Query Match 5.0%; Score 8; DB 2; Length 539;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 137 DVSARDVT 144
||||||||
Db 4 DVSARDVT 11

RESULT 9
A84833
hypothetical protein At2g40720 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84833
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84833
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-860 <STO>
A;Cross-references: GB:AE002093; NID:g4895221; PIDN:AAD32807.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g40720
A;Map position: 2

Query Match 5.0%; Score 8; DB 2; Length 860;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 138 VSARDVTV 145
||||||||
Db 128 VSARDVTV 135

RESULT 10
T30160
hypothetical protein C37A2.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T30160
R;Le, T.T.; Kemp, K.; Scheet, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid C37A2.
A;Reference number: Z20746
A;Accession: T30160
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1234 <LET>
A;Cross-references: EMBL:U97194; PIDN:AAB52447.1; GSPDB:GN00019; CESP:C37A2.2
A;Experimental source: strain Bristol N2; clone C37A2
C;Genetics:
A;Gene: CESP:C37A2.2
A;Map position: 1
A;Introns: 46/1; 124/1; 159/1; 231/1; 508/3; 740/3; 891/1; 931/1; 975/1; 1014/3; 1083/3

Query Match 5.0%; Score 8; DB 2; Length 1234;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 RGSVPIPL 158
||||||||
Db 960 RGSVPIPL 967

RESULT 11
T30944
surface protein precursor - Enterococcus faecalis
C;Species: Enterococcus faecalis
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30944
R;Shankar, V.; Baghdayan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
Infect. Immun. 67, 193-200, 1999
A;Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene
A;Reference number: Z20943; MUID:99081742; PMID:9864215
A;Accession: T30944
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1873 <SHA>
A;Cross-references: EMBL:AF034779; NID:g3873186; PID:g3873187; PIDN:AAD09858.1

Query Match 5.0%; Score 8; DB 2; Length 1873;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 DVTVTLPD 149
||||||||
Db 1754 DVTVTLPD 1761

RESULT 12
B69932
hypothetical protein yozN - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: B69932
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69932
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-87 <KUN>
A;Cross-references: GB:Z99114; GB:AL009126; NID:g2634230; PIDN:CAB13819.1; PID:e11853
A;Experimental source: strain 168
C;Genetics:
A;Gene: yozN

Query Match 4.3%; Score 7; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 VNVGQNL 9
|||||||
Db 26 VNVGQNL 32

RESULT 13
H96717
probable 60S ribosomal protein L34 T6C23.18 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: H96717
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96717
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-119 <STO>
A;Cross-references: GB:AE005173; NID:g6665554; PIDN:AAF22923.1; GSPDB:GN00141
C;Genetics:
A;Gene: T6C23.18
A;Map position: 1
C;Superfamily: rat ribosomal protein L34

Query Match 4.3%; Score 7; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 AYGGVLS 44
|||||||
Db 77 AYGGVLS 83

RESULT 14
S48027
ribosomal protein L34, cytosolic - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 13-Aug-1999
C;Accession: S48027; S48028
R;Gao, J.; Kim, S.R.; Chung, Y.Y.; Lee, J.M.; An, G.
Plant Mol. Biol. 25, 761-770, 1994
A;Title: Developmental and environmental regulation of two ribosomal protein genes in to
A;Reference number: S48026; MUID:94355650; PMID:8075394
A;Accession: S48027
A;Molecule type: mRNA
A;Residues: 1-120 <GAO>
A;Cross-references: EMBL:L27089; NID:g436029; PIDN:AAA57158.1; PID:g436030
A;Experimental source: TSC40-3
A;Accession: S48028
A;Molecule type: mRNA
A;Residues: 1-120 <GA2>
A;Cross-references: EMBL:L27107; NID:g436031; PIDN:AAA57159.1; PID:g436032
A;Experimental source: TSC40-4
C;Superfamily: rat ribosomal protein L34
C;Keywords: protein biosynthesis; ribosome

Query Match 4.3%; Score 7; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 AYGGVLS 44
|||||||
Db 77 AYGGVLS 83

RESULT 15
S60476
ribosomal protein L34, cytosolic - garden pea
C;Species: Pisum sativum (garden pea)
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 13-Aug-1999
C;Accession: S60476
R;Devitt, M.L.; Stafstrom, J.P.
Plant Mol. Biol. 29, 255-265, 1995
A;Title: Cell cycle regulation during growth-dormancy cycles in pea axillary buds.
A;Reference number: S60474; MUID:96046745; PMID:7579177
A;Accession: S60476
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-120 <DEV>
A;Cross-references: EMBL:U10047; NID:g498907; PIDN:AAA86953.1; PID:g498908
C;Genetics:
A;Gene: rpl34
C;Superfamily: rat ribosomal protein L34
C;Keywords: protein biosynthesis; ribosome

Query Match 4.3%; Score 7; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 AYGGVLS 44
|||||||
Db 77 AYGGVLS 83

RESULT 16
F86395
60s ribosomal protein L34 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: F86395
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F86395
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-120 <STO>
A;Cross-references: GB:AE005172; NID:g4262177; PIDN:AAD14494.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: rat ribosomal protein L34

Query Match 4.3%; Score 7; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 AYGGVLS 44
|||||||
Db 77 AYGGVLS 83

RESULT 17
B75531
hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C;Accession: B75531
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: B75531
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-128 <WHI>
A;Cross-references: GB:AE001894; GB:AE000513; NID:g6458011; PIDN:AAF09928.1; PID:g645

```
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0338
A;Map position: 1
C;Superfamily: Deinococcus radiodurans hypothetical protein DR0338

  Query Match             4.3%;  Score 7;  DB 2;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      43 LSNFSGT 49
           |||||||
Db      23 LSNFSGT 29


RESULT 18
MPHU2
myelin P2 protein [validated] - human
N;Alternate names: peripheral myelin protein 2
C;Species: Homo sapiens (man)
C;Date: 05-Apr-1983 #sequence_revision 27-Jan-1995 #text_change 08-Dec-2000
C;Accession: JT0977; A03143; S24224
R;Hayasaka, K.; Nanao, K.; Tahara, M.; Sato, W.; Takada, G.; Miura, M.; Uyemura, K.
Biochem. Biophys. Res. Commun. 181, 204-207, 1991
A;Title: Isolation and sequence determination of cDNA encoding P2 protein of human perip
A;Reference number: JT0977; MUID:92068191; PMID:1720307
A;Accession: JT0977
A;Molecule type: mRNA
A;Residues: 1-132 <HAY>
A;Cross-references: EMBL:X62167; NID:g35185; PIDN:CAA44096.1; PID:g35186
A;Experimental source: fetal spinal cord peripheral myelin
A;Note: authors did not translate the codon for residue 1
R;Suzuki, M.; Kitamura, K.; Sakamoto, Y.; Uyemura, K.
J. Neurochem. 39, 1759-1762, 1982
A;Title: The complete amino acid sequence of human P2 protein.
A;Reference number: A03143; MUID:83058785; PMID:6183401
A;Accession: A03143
A;Molecule type: protein
A;Residues: 2-98,'N',100-110,'D',112-132 <SUZ>
C;Comment: P2 protein and myelin basic protein together constitute a major fraction of p
C;Genetics:
A;Gene: GDB:PMP2
A;Cross-references: GDB:129030; OMIM:170715
A;Map position: 8q21.3-8q22.1
C;Superfamily: myelin P2 protein
C;Keywords: acetylated amino end; myelin; phosphoprotein
F;2-132/Product: myelin P2 protein #status experimental <MAT>
F;2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F;20/Binding site: phosphate (Tyr) (covalent) #status predicted
F;118-125/Disulfide bonds: #status experimental

  Query Match             4.3%;  Score 7;  DB 1;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 VTLQRGS 37
           |||||||
Db      85 VTLQRGS 91


RESULT 19
A98226
hypothetical protein AGR_L_1506 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: A98226
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: A98226
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-162 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89331.1; PID:g15159173; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1506
A;Map position: linear chromosome

  Query Match             4.3%;  Score 7;  DB 2;  Length 162;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 NVGQNLV 10
           |||||||
Db      56 NVGQNLV 62


RESULT 20
AG3060
conserved hypothetical protein Atu4100 [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AG3060
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AG3060
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-162 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44901.1; PID:g17742552; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu4100
A;Map position: linear chromosome

  Query Match             4.3%;  Score 7;  DB 2;  Length 162;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 NVGQNLV 10
           |||||||
Db      56 NVGQNLV 62


RESULT 21
B82696
cytochrome B561 XF1328 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: B82696
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: B82696
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-188 <SIM>
A;Cross-references: GB:AE003965; GB:AE003849; NID:g9106313; PIDN:AAF84137.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
```

A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1328
C;Superfamily: cytochrome b561

Query Match 4.3%; Score 7; DB 2; Length 188;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 LIAVLIL 106
||||||| |
Db 20 LIAVLIL 26

RESULT 22
D86724
hypothetical protein yiaD [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: D86724
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: D86724
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-210 <STO>
A;Cross-references: GB:AE005176; PID:g12723717; PIDN:AAK04894.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yiaD

Query Match 4.3%; Score 7; DB 2; Length 210;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 96 KAGSLIA 102
||||||| |
Db 12 KAGSLIA 18

RESULT 23
E70600
hypothetical protein Rv3908 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 26-May-2000
C;Accession: E70600
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70600
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-248 <COL>
A;Cross-references: GB:Z94121; GB:AL123456; NID:g3261736; PIDN:CAB08093.1; PID:e312269;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3908
C;Superfamily: unassigned mutT domain proteins; mutT domain homology
F;98-135/Domain: mutT domain homology <MUTT>

Query Match 4.3%; Score 7; DB 2; Length 248;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 SAGGLVI 95
||||||| |
Db 66 SAGGLVI 72

RESULT 24
H87246
conserved hypothetical protein ML2698 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 24-Aug-2001
C;Accession: H87246
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.,
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.,
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: H87246
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-251 <STO>
A;Cross-references: GB:AL450380; NID:g13093888; PIDN:CAC32230.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML2698
C;Superfamily: unassigned mutT domain proteins; mutT domain homology

Query Match 4.3%; Score 7; DB 2; Length 251;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 SAGGLVI 95
||||||| |
Db 69 SAGGLVI 75

RESULT 25
F69959
hypothetical protein yqhO - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: F69959
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F69959
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-291 <KUN>
A;Cross-references: GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14382.1; PID:e11857
A;Experimental source: strain 168
C;Genetics:
A;Gene: yqhO

Query Match 4.3%; Score 7; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      40 GGVLSNF 46
           |||||||
Db     184 GGVLSNF 190
```

RESULT 26
S24169
mucin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 10-Nov-1995
C;Accession: S24169
R;Huan, L.J.; Xu, G.; Forstner, G.; Forstner, J.
Biochim. Biophys. Acta 1132, 79-82, 1992
A;Title: A serine, threonine and proline-rich region near the carboxyl-terminus of a rat
A;Reference number: S24169; MUID:92379096; PMID:1380835
A;Accession: S24169
A;Molecule type: mRNA
A;Residues: 1-292 <HUA>

```
  Query Match             4.3%;  Score 7;  DB 2;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 PTTSETP 66
           |||||||
Db       7 PTTSETP 13
```

RESULT 27
E97096
Zn-binding lipoprotein related (surface adhesin A), ADHS [imported] - Clostridium acetob
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 01-Mar-2002
C;Accession: E97096
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: E97096
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-298 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK79560.1; PID:g15024548; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1593
C;Superfamily: adhesin B

```
  Query Match             4.3%;  Score 7;  DB 2;  Length 298;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 SLIAVLI 105
           |||||||
Db       7 SLIAVLI 13
```

RESULT 28
H75081
phosphate abc transporter, permease protein (pstc) PAB0698 - Pyrococcus abyssi (strain C
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: H75081
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: H75081
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-304 <KAW>
A;Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB49957.1; PID:g545
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: pstC; PAB0698
C;Superfamily: phoW protein

```
  Query Match             4.3%;  Score 7;  DB 2;  Length 304;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 SLIAVLI 105
           |||||||
Db      72 SLIAVLI 78
```

RESULT 29
T32776
hypothetical protein D1069.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
C;Accession: T32776
R;Murray, J.; Langston, Y.; Clarke, K.; Morris, M.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid D1069.
A;Reference number: Z21223
A;Accession: T32776
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-311 <MUR>
A;Cross-references: EMBL:AF040641; PIDN:AAB94946.1; GSPDB:GN00020; CESP:D1069.1
A;Experimental source: strain Bristol N2; clone D1069
C;Genetics:
A;Gene: CESP:D1069.1
A;Map position: 2
A;Introns: 264/3
C;Superfamily: Caenorhabditis elegans hypothetical protein D1069.1

```
  Query Match             4.3%;  Score 7;  DB 2;  Length 311;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 VVDLSTQ 16
           |||||||
Db     211 VVDLSTQ 217
```

RESULT 30
T38762
hypothetical protein SPAC3H8.04 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
C;Accession: T38762
R;Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z21810
A;Accession: T38762
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-338 <GEN>
A;Cross-references: EMBL:Z69086; PIDN:CAA93161.1; GSPDB:GN00066; SPDB:SPAC3H8.04
A;Experimental source: strain 972h-; cosmid c3H8
C;Genetics:
A;Gene: SPDB:SPAC3H8.04
A;Map position: 1
C;Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC3H8.04

```
  Query Match             4.3%;  Score 7;  DB 2;  Length 338;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     104 LILRQTN 110
           |||||||
Db     237 LILRQTN 243
```

RESULT 31
AD0183
probable exported protein YPO1504 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AD0183
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AD0183
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-383 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC90327.1; PID:g15979546; GSPDB:GN00175
C;Genetics:
A;Gene: YPO1504

Query Match 4.3%; Score 7; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 AGGLVIK 96
|||||||
Db 366 AGGLVIK 372

RESULT 32
AG2704
MFS permease [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 02-Aug-2002
C;Accession: AG2704
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AG2704
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-406 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL42053.1; PID:g17739431; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu1040
A;Map position: circular chromosome
C;Superfamily: hypothetical protein c0103

Query Match 4.3%; Score 7; DB 2; Length 406;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 AGSLIAV 103
|||||||
Db 18 AGSLIAV 24

RESULT 33
G97486
probable mfS transporter (PA1286) [imported] - Agrobacterium tumefaciens (strain C58, Ce
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 02-Aug-2002
C;Accession: G97486
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A;Reference number: A97359; PMID:11743194
A;Accession: G97486
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-414 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK86848.1; PID:g15156062; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_1918
A;Map position: circular chromosome
C;Superfamily: hypothetical protein c0103

Query Match 4.3%; Score 7; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 AGSLIAV 103
|||||||
Db 26 AGSLIAV 32

RESULT 34
B86849
polysacharide biosynthesis protein [imported] - Lactococcus lactis subsp. lactis (str
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: B86849
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: B86849
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-443 <STO>
A;Cross-references: GB:AE005176; PID:g12724819; PIDN:AAK05892.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: ysfC

Query Match 4.3%; Score 7; DB 2; Length 443;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 SAGGLVI 95
|||||||
Db 57 SAGGLVI 63

RESULT 35
F98149
hypothetical protein AGR_L_305 [imported] - Agrobacterium tumefaciens (strain C58, Ce
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: F98149
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A;Reference number: A97359; PMID:11743194
A;Accession: F98149
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-498 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK88720.1; PID:g15158457; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_305
A;Map position: linear chromosome
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog

Query Match 4.3%; Score 7; DB 2; Length 498;
Best Local Similarity 100.0%; Pred. No. 56;

```
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     142 DVTVTLP 148
           |||||||
Db     285 DVTVTLP 291


RESULT 36
AE3138
hypothetical protein Atu4729 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C;Accession: AE3138
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AE3138
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-533 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL45523.1; PID:g17743234; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu4729
A;Map position: linear chromosome
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

  Query Match             4.3%;  Score 7;  DB 2;  Length 533;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     142 DVTVTLP 148
           |||||||
Db     320 DVTVTLP 326


RESULT 37
F75501
sensor histidine kinase - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: F75501
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75501
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-590 <WHI>
A;Cross-references: GB:AE001917; GB:AE000513; NID:g6458281; PIDN:AAF10157.1; PID:g645827
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0577
A;Map position: 1

  Query Match             4.3%;  Score 7;  DB 2;  Length 590;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      87 VSSAGGL 93
           |||||||
Db     549 VSSAGGL 555


RESULT 38
S29685
retroviral receptor - human
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Nov-2000
C;Accession: S29685; A40775
R;Abritton, M.; Tseng, L.; Cunningham, J.M.
submitted to the EMBL Data Library, January 1991
A;Description: Nucleotide sequence of the human gene similar to the murine ecotropic
A;Reference number: S29685
A;Accession: S29685
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-629 <ABR>
A;Cross-references: EMBL:X57303; NID:g35919; PIDN:CAA40560.1; PID:g35920
R;Yoshimoto, T.; Yoshimoto, E.; Meruelo, D.
Virology 185, 10-17, 1991
A;Title: Molecular cloning and characterization of a novel human gene homologous to t
A;Reference number: A40775; MUID:92024065; PMID:1718082
A;Accession: A40775
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-22,'R',24-629 <YOS>
A;Cross-references: GB:X59155; NID:g36160; PIDN:CAA41869.1; PID:g36161
C;Genetics:
A;Gene: REC1L
C;Superfamily: ecotropic retrovirus receptor protein

  Query Match             4.3%;  Score 7;  DB 2;  Length 629;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 SLIAVLI 105
           |||||||
Db     501 SLIAVLI 507


RESULT 39
G84334
threonyl-tRNA synthetase [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: G84334
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: G84334
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-650 <STO>
A;Cross-references: GB:AE004437; NID:g10581281; PIDN:AAG20043.1; GSPDB:GN00138
C;Genetics:
A;Gene: thrS
C;Superfamily: threonine-tRNA ligase

  Query Match             4.3%;  Score 7;  DB 2;  Length 650;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     143 VTVTLPD 149
           |||||||
Db      14 VTVTLPD 20


RESULT 40
S74621
carbon dioxide concentrating mechanism protein ccmM - Synechocystis sp. (strain PCC 6
N;Alternate names: hypothetical protein sll1031
C;Species: Synechocystis sp.
A;Variety: PCC 6803
```

C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S74621
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S74621
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-687 <KAN>
A;Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BAA16773.1; PID:d101750
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: ccmM
A;Start codon: GTG

Query Match 4.3%; Score 7; DB 2; Length 687;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 PVSSAGG 92
|||||||
Db 466 PVSSAGG 472

RESULT 41
D81798
lactoferrin-binding protein NMA1740 [imported] - Neisseria meningitidis (strain Z2491 se
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: D81798
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: D81798
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-741 <PAR>
A;Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB84968.1; PID:g738038
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: lbpB; NMA1740

Query Match 4.3%; Score 7; DB 2; Length 741;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 GTVKYSG 54
|||||||
Db 194 GTVKYSG 200

RESULT 42
G90728
probable enzyme [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 24-Aug-2001
C;Accession: G90728
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G90728
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-761 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34222.1; PID:g13360258; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0799
C;Superfamily: iron-responsive element-binding protein

Query Match 4.3%; Score 7; DB 2; Length 761;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 IKAGSLI 101
|||||||
Db 747 IKAGSLI 753

RESULT 43
H85579
probable enzyme ybhJ [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: H85579
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: H85579
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-761 <STO>
A;Cross-references: GB:AE005174; NID:g12513706; PIDN:AAG55100.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: ybhJ
C;Superfamily: iron-responsive element-binding protein

Query Match 4.3%; Score 7; DB 2; Length 761;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 IKAGSLI 101
|||||||
Db 747 IKAGSLI 753

RESULT 44
C64813
ybhJ protein - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: C64813
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: C64813
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-761 <BLAT>
A;Cross-references: GB:AE000179; GB:U00096; NID:g1786978; PIDN:AAC73858.1; PID:g17869
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: ybhJ
C;Superfamily: iron-responsive element-binding protein
C;Keywords: transmembrane protein
F;413-429/Domain: transmembrane #status predicted <TMM>

Query Match 4.3%; Score 7; DB 2; Length 761;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 IKAGSLI 101
|||||||
Db 747 IKAGSLI 753

RESULT 45
T30081
hypothetical protein C31H2.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T30081
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C31H2.
A;Reference number: Z20732
A;Accession: T30081
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-790 <GEI>
A;Cross-references: EMBL:U41748; PIDN:AAA83335.1; CESP:C31H2.1
C;Genetics:
A;Gene: CESP:C31H2.1
A;Introns: 45/3; 78/2; 110/3; 148/3; 198/1; 244/3; 274/3; 318/3; 366/3; 396/3; 427/3; 45
C;Superfamily: Caenorhabditis elegans hypothetical protein C31H2.1

Query Match 4.3%; Score 7; DB 2; Length 790;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 40 GGVLSNF 46
|||||||
Db 210 GGVLSNF 216

Search completed: November 28, 2002, 19:06:08
Job time : 32 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 19:01:31 ; Search time 17 Seconds
(without alignments)
278.652 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 161
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	4.3	132	1	US-08-409-731A-11	Sequence 11, Appl
2	7	4.3	132	2	US-08-470-298B-11	Sequence 11, Appl
3	7	4.3	132	2	US-09-023-073A-11	Sequence 11, Appl
4	7	4.3	132	4	US-09-361-737-11	Sequence 11, Appl
5	7	4.3	629	2	US-08-132-990A-8	Sequence 8, Appli
6	7	4.3	629	5	PCT-US92-09382-8	Sequence 8, Appli
7	6	3.7	92	4	US-09-134-001C-4014	Sequence 4014, Ap
8	6	3.7	124	1	US-08-326-117B-13	Sequence 13, Appl
9	6	3.7	124	3	US-08-982-129-13	Sequence 13, Appl
10	6	3.7	124	4	US-08-983-607-51	Sequence 51, Appl
11	6	3.7	146	4	US-09-134-001C-4330	Sequence 4330, Ap
12	6	3.7	185	3	US-08-985-526-36	Sequence 36, Appl
13	6	3.7	191	4	US-09-561-500-13	Sequence 13, Appl
14	6	3.7	191	4	US-09-561-108-13	Sequence 13, Appl
15	6	3.7	191	4	US-09-561-526-13	Sequence 13, Appl
16	6	3.7	195	1	US-08-159-784-2	Sequence 2, Appli
17	6	3.7	213	2	US-08-248-839C-94	Sequence 94, Appl
18	6	3.7	243	2	US-07-885-089B-4	Sequence 4, Appli
19	6	3.7	243	2	US-07-885-089B-9	Sequence 9, Appli
20	6	3.7	243	4	US-08-944-483-70	Sequence 70, Appl
21	6	3.7	243	6	5202428-2	Patent No. 5202428
22	6	3.7	247	2	US-08-956-267A-2	Sequence 2, Appli
23	6	3.7	249	2	US-08-989-386-5	Sequence 5, Appli
24	6	3.7	252	4	US-09-238-303-10	Sequence 10, Appl
25	6	3.7	262	4	US-08-491-944-2	Sequence 2, Appli
26	6	3.7	281	4	US-09-025-769B-178	Sequence 178, App
27	6	3.7	293	4	US-09-512-342-14	Sequence 14, Appl
28	6	3.7	297	1	US-08-118-270-57	Sequence 57, Appl
29	6	3.7	297	1	US-08-118-270-58	Sequence 58, Appl
30	6	3.7	297	5	PCT-US93-08528-57	Sequence 57, Appl
31	6	3.7	297	5	PCT-US93-08528-58	Sequence 58, Appl
32	6	3.7	307	4	US-09-134-001C-3238	Sequence 3238, Ap
33	6	3.7	329	4	US-09-424-349A-8	Sequence 8, Appli
34	6	3.7	347	2	US-08-164-292B-13	Sequence 13, Appl
35	6	3.7	347	3	US-08-845-623-13	Sequence 13, Appl
36	6	3.7	347	3	US-08-815-927-13	Sequence 13, Appl
37	6	3.7	347	4	US-09-103-330-13	Sequence 13, Appl
38	6	3.7	347	4	US-09-435-242-13	Sequence 13, Appl
39	6	3.7	358	4	US-09-615-192A-338	Sequence 338, App
40	6	3.7	366	4	US-09-134-001C-4365	Sequence 4365, Ap
41	6	3.7	376	4	US-09-134-001C-4259	Sequence 4259, Ap
42	6	3.7	387	3	US-08-689-421-23	Sequence 23, Appl
43	6	3.7	387	4	US-09-389-528-23	Sequence 23, Appl
44	6	3.7	387	4	US-09-181-827A-23	Sequence 23, Appl
45	6	3.7	391	4	US-09-131-028A-4	Sequence 4, Appli
46	6	3.7	391	4	US-09-131-028A-14	Sequence 14, Appl
47	6	3.7	417	4	US-08-867-611-20	Sequence 20, Appl
48	6	3.7	417	5	PCT-US92-06965A-25	Sequence 25, Appl
49	6	3.7	445	2	US-08-630-118A-2	Sequence 2, Appli
50	6	3.7	445	2	US-08-630-118A-4	Sequence 4, Appli
51	6	3.7	445	2	US-08-838-399-2	Sequence 2, Appli
52	6	3.7	445	2	US-08-838-399-4	Sequence 4, Appli
53	6	3.7	445	2	US-09-003-199-2	Sequence 2, Appli
54	6	3.7	445	2	US-09-003-199-23	Sequence 23, Appl
55	6	3.7	445	4	US-09-235-839-2	Sequence 2, Appli
56	6	3.7	445	4	US-09-235-839-4	Sequence 4, Appli
57	6	3.7	445	4	US-09-327-035-2	Sequence 2, Appli
58	6	3.7	445	4	US-09-327-035-4	Sequence 4, Appli
59	6	3.7	447	4	US-09-370-253-6	Sequence 6, Appli
60	6	3.7	453	3	US-08-993-359-22	Sequence 22, Appl
61	6	3.7	455	2	US-08-272-255-14	Sequence 14, Appl
62	6	3.7	455	5	PCT-US95-08565-14	Sequence 14, Appl
63	6	3.7	456	1	US-08-349-025-2	Sequence 2, Appli
64	6	3.7	456	2	US-08-566-096A-2	Sequence 2, Appli
65	6	3.7	456	2	US-08-668-650B-2	Sequence 2, Appli
66	6	3.7	456	4	US-09-200-673-2	Sequence 2, Appli
67	6	3.7	456	5	PCT-US95-15646-2	Sequence 2, Appli
68	6	3.7	466	4	US-09-134-001C-3355	Sequence 3355, Ap
69	6	3.7	487	4	US-09-525-046-2	Sequence 2, Appli
70	6	3.7	516	3	US-08-689-421-29	Sequence 29, Appl
71	6	3.7	516	4	US-09-389-528-29	Sequence 29, Appl
72	6	3.7	516	4	US-09-181-827A-29	Sequence 29, Appl
73	6	3.7	521	3	US-08-996-338-20	Sequence 20, Appl
74	6	3.7	530	3	US-09-222-817-12	Sequence 12, Appl
75	6	3.7	530	3	US-09-222-817-14	Sequence 14, Appl
76	6	3.7	530	4	US-09-222-786-12	Sequence 12, Appl
77	6	3.7	530	4	US-09-222-786-14	Sequence 14, Appl
78	6	3.7	531	3	US-08-688-988-39	Sequence 39, Appl
79	6	3.7	541	1	US-08-604-333-2	Sequence 2, Appli
80	6	3.7	541	3	US-09-110-618-2	Sequence 2, Appli
81	6	3.7	541	4	US-09-173-151A-28	Sequence 28, Appl
82	6	3.7	541	4	US-09-578-178-2	Sequence 2, Appli
83	6	3.7	551	4	US-09-130-491-16	Sequence 16, Appl
84	6	3.7	576	1	US-08-221-817-13	Sequence 13, Appl
85	6	3.7	576	1	US-08-454-439-13	Sequence 13, Appl
86	6	3.7	576	4	US-08-464-954A-6	Sequence 6, Appli
87	6	3.7	576	5	PCT-US94-10487-13	Sequence 13, Appl
88	6	3.7	579	1	US-08-126-564A-31	Sequence 31, Appl
89	6	3.7	579	5	PCT-US94-09143-31	Sequence 31, Appl
90	6	3.7	596	4	US-09-171-337A-6	Sequence 6, Appli
91	6	3.7	632	1	US-08-221-817-11	Sequence 11, Appl
92	6	3.7	632	1	US-08-454-439-11	Sequence 11, Appl
93	6	3.7	632	5	PCT-US94-10487-11	Sequence 11, Appl
94	6	3.7	637	4	US-09-221-294-4	Sequence 4, Appli
95	6	3.7	640	4	US-09-351-814-13	Sequence 13, Appl
96	6	3.7	667	4	US-08-959-004-11	Sequence 11, Appl
97	6	3.7	675	4	US-08-973-273-5	Sequence 5, Appli
98	6	3.7	726	4	US-09-413-814-30	Sequence 30, Appl
99	6	3.7	808	2	US-08-629-291A-33	Sequence 33, Appl
100	6	3.7	808	2	US-08-658-335B-33	Sequence 33, Appl

101	6	3.7	808	4	US-09-406-640-33
102	6	3.7	897	4	US-09-134-001C-3600
103	6	3.7	921	1	US-08-188-582-2
104	6	3.7	921	1	US-08-646-715-2
105	6	3.7	1027	2	US-08-551-437-2
106	6	3.7	1027	3	US-09-004-225-2
107	6	3.7	1027	4	US-09-084-346-2
108	6	3.7	1027	4	US-09-104-704-2
109	6	3.7	1375	4	US-09-171-410-1
110	6	3.7	1404	4	US-08-801-308-1
111	6	3.7	1528	1	US-08-326-117B-2
112	6	3.7	1528	3	US-08-982-129-2
113	6	3.7	1708	1	US-08-493-092-2
114	6	3.7	1708	1	US-08-508-836A-2
115	6	3.7	1717	4	US-09-178-176B-2
116	6	3.7	1717	4	US-09-457-864-2
117	6	3.7	2254	2	US-08-286-819A-28
118	6	3.7	2254	3	US-08-980-357-28
119	6	3.7	2496	4	US-09-125-028-2
120	6	3.7	2958	4	US-08-894-344C-2
121	6	3.7	3056	1	US-08-508-836A-8
122	6	3.7	3056	2	US-08-629-001A-3
123	6	3.7	3056	2	US-08-874-266-2
124	6	3.7	3056	4	US-08-642-274D-3
125	6	3.7	3056	4	US-08-952-127-3
126	6	3.7	3056	4	US-08-952-014C-3
127	6	3.7	3056	4	US-09-360-416-2
128	6	3.7	3057	4	US-09-360-416-3
129	6	3.7	3066	4	US-08-952-127-12
130	6	3.7	5215	4	US-09-105-537-2
131	6	3.7	10182	4	US-09-134-001C-3159
132	5	3.1	5	3	US-08-894-173-35
133	5	3.1	5	4	US-09-398-193-35
134	5	3.1	7	2	US-09-070-756-7
135	5	3.1	8	3	US-08-747-221B-43
136	5	3.1	8	4	US-09-005-051-43
137	5	3.1	10	4	US-09-155-427-4
138	5	3.1	11	4	US-09-425-638A-72
139	5	3.1	11	4	US-09-543-004-72
140	5	3.1	12	3	US-08-747-221B-39
141	5	3.1	12	3	US-08-747-221B-41
142	5	3.1	12	4	US-09-005-051-39
143	5	3.1	12	4	US-09-005-051-41
144	5	3.1	13	2	US-08-480-190-229
145	5	3.1	13	2	US-08-488-379-229
146	5	3.1	13	4	US-09-385-442-44
147	5	3.1	13	4	US-09-252-586-25
148	5	3.1	13	4	US-09-259-478A-22
149	5	3.1	13	5	PCT-US93-07545-229
150	5	3.1	14	2	US-08-574-959A-12
151	5	3.1	14	4	US-09-357-014-12
152	5	3.1	14	6	5190919-31
153	5	3.1	15	2	US-08-676-279-7
154	5	3.1	15	2	US-07-728-215-43
155	5	3.1	15	4	US-08-743-168B-22
156	5	3.1	15	4	US-08-938-085A-49
157	5	3.1	15	5	PCT-US96-10435-22
158	5	3.1	16	6	5245013-10
159	5	3.1	17	5	PCT-US95-08621-25
160	5	3.1	18	4	US-09-259-478A-15
161	5	3.1	20	1	US-08-468-543-23
162	5	3.1	20	1	US-08-484-135-63
163	5	3.1	20	1	US-08-484-635-233
164	5	3.1	20	2	US-08-484-631-233
165	5	3.1	20	2	US-08-469-692-23
166	5	3.1	20	2	US-08-398-046-23
167	5	3.1	20	2	US-08-827-570-233
168	5	3.1	21	3	US-08-747-221B-42
169	5	3.1	21	4	US-09-005-051-42
170	5	3.1	21	6	5190919-4
171	5	3.1	22	1	US-07-698-928A-1
172	5	3.1	22	1	US-07-698-928A-5
173	5	3.1	22	3	US-08-747-221B-40

Sequence 33, Appl
Sequence 3600, Ap
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 28, Appl
Sequence 28, Appl
Sequence 2, Appli
Sequence 2, Appli
Sequence 8, Appli
Sequence 3, Appli
Sequence 2, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 2, Appli
Sequence 3, Appli
Sequence 12, Appl
Sequence 2, Appli
Sequence 3159, Ap
Sequence 35, Appl
Sequence 35, Appl
Sequence 7, Appli
Sequence 43, Appl
Sequence 43, Appl
Sequence 4, Appli
Sequence 72, Appl
Sequence 72, Appl
Sequence 39, Appl
Sequence 41, Appl
Sequence 39, Appl
Sequence 41, Appl
Sequence 229, App
Sequence 229, App
Sequence 44, Appl
Sequence 25, Appl
Sequence 22, Appl
Sequence 229, App
Sequence 12, Appl
Sequence 12, Appl
Patent No. 5190919
Sequence 7, Appli
Sequence 43, Appl
Sequence 22, Appl
Sequence 49, Appl
Sequence 22, Appl
Patent No. 5245013
Sequence 25, Appl
Sequence 15, Appl
Sequence 23, Appl
Sequence 63, Appl
Sequence 233, App
Sequence 233, App
Sequence 23, Appl
Sequence 23, Appl
Sequence 233, App
Sequence 42, Appl
Sequence 42, Appl
Patent No. 5190919
Sequence 1, Appli
Sequence 5, Appli
Sequence 40, Appl

174	5	3.1	22	4	US-09-005-051-40
175	5	3.1	22	4	US-09-028-937-22
176	5	3.1	23	1	US-08-383-753-79
177	5	3.1	23	2	US-08-586-772-79
178	5	3.1	23	2	US-08-637-759B-299
179	5	3.1	23	2	US-08-959-512-79
180	5	3.1	23	3	US-08-871-355A-299
181	5	3.1	23	3	US-08-747-221B-44
182	5	3.1	23	4	US-09-512-983-79
183	5	3.1	23	4	US-09-005-051-44
184	5	3.1	23	4	US-09-201-945-299
185	5	3.1	24	1	US-07-698-928A-8
186	5	3.1	24	3	US-08-782-997A-35
187	5	3.1	24	4	US-08-913-942-13
188	5	3.1	26	1	US-07-942-245-364
189	5	3.1	26	1	US-07-942-245-408
190	5	3.1	26	1	US-08-468-543-24
191	5	3.1	26	2	US-08-469-692-24
192	5	3.1	26	2	US-08-398-046-24
193	5	3.1	27	1	US-08-419-594-2
194	5	3.1	27	2	US-08-919-089-2
195	5	3.1	27	3	US-08-856-383-5
196	5	3.1	27	4	US-09-475-088-5
197	5	3.1	28	4	US-08-469-260A-573
198	5	3.1	29	3	US-09-045-632-98
199	5	3.1	29	4	US-09-060-756-732
200	5	3.1	29	6	5514646-49
201	5	3.1	30	4	US-08-981-392-50
202	5	3.1	33	2	US-08-671-487A-1
203	5	3.1	33	3	US-08-476-376-5
204	5	3.1	34	4	US-08-734-607B-5
205	5	3.1	34	4	US-08-734-607B-13
206	5	3.1	35	6	5190919-5
207	5	3.1	35	6	5190919-6
208	5	3.1	36	1	US-08-475-989-7
209	5	3.1	36	2	US-08-475-985-7
210	5	3.1	36	2	US-08-472-172-19
211	5	3.1	36	3	US-08-256-839-7
212	5	3.1	36	6	5190919-15
213	5	3.1	37	4	US-09-012-669F-31
214	5	3.1	38	3	US-08-545-860D-67
215	5	3.1	38	5	PCT-US94-04496-67
216	5	3.1	39	3	US-08-476-376-7
217	5	3.1	39	6	5245013-1
218	5	3.1	41	4	US-09-227-357-670
219	5	3.1	43	4	US-09-100-193-10
220	5	3.1	43	4	US-09-202-316-12
221	5	3.1	44	4	US-09-155-427-2
222	5	3.1	46	4	US-08-857-076-83
223	5	3.1	47	2	US-08-676-279-54
224	5	3.1	48	4	US-09-071-035-196
225	5	3.1	50	4	US-09-007-741-1
226	5	3.1	52	4	US-09-346-860-1
227	5	3.1	54	1	US-08-439-546-1
228	5	3.1	54	1	US-08-439-348-1
229	5	3.1	54	2	US-08-601-478-1
230	5	3.1	54	3	US-08-750-419A-10
231	5	3.1	55	2	US-08-102-385G-16
232	5	3.1	56	1	US-08-328-256-9
233	5	3.1	57	1	US-07-956-700B-74
234	5	3.1	57	1	US-08-476-537-74
235	5	3.1	57	1	US-08-485-607-74
236	5	3.1	57	2	US-08-475-879-74
237	5	3.1	57	4	US-09-433-043B-74
238	5	3.1	58	4	US-08-734-607B-17
239	5	3.1	59	1	US-08-470-179-6
240	5	3.1	59	2	US-08-933-750C-3
241	5	3.1	59	4	US-09-234-613-3
242	5	3.1	60	3	US-08-468-011A-20
243	5	3.1	60	4	US-09-188-930-181
244	5	3.1	60	4	US-09-188-930-320
245	5	3.1	60	4	US-09-236-468A-20
246	5	3.1	63	4	US-08-963-851-18

Sequence 40, Appl
Sequence 22, Appl
Sequence 79, Appl
Sequence 79, Appl
Sequence 299, App
Sequence 79, Appl
Sequence 299, App
Sequence 44, Appl
Sequence 79, Appl
Sequence 44, Appl
Sequence 299, App
Sequence 8, Appli
Sequence 35, Appl
Sequence 13, Appl
Sequence 364, App
Sequence 408, App
Sequence 24, Appl
Sequence 24, Appl
Sequence 24, Appl
Sequence 2, Appli
Sequence 2, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 573, App
Sequence 98, Appl
Sequence 732, App
Patent No. 5514646
Sequence 50, Appl
Sequence 1, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 13, Appl
Patent No. 5190919
Patent No. 5190919
Sequence 7, Appli
Sequence 7, Appli
Sequence 19, Appl
Sequence 7, Appli
Patent No. 5190919
Sequence 31, Appl
Sequence 67, Appl
Sequence 67, Appl
Sequence 7, Appli
Patent No. 5245013
Sequence 670, App
Sequence 10, Appl
Sequence 12, Appl
Sequence 2, Appli
Sequence 83, Appl
Sequence 54, Appl
Sequence 196, App
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 10, Appl
Sequence 16, Appl
Sequence 9, Appli
Sequence 74, Appl
Sequence 74, Appl
Sequence 74, Appl
Sequence 74, Appl
Sequence 74, Appl
Sequence 17, Appl
Sequence 6, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 20, Appl
Sequence 181, App
Sequence 320, App
Sequence 20, Appl
Sequence 18, Appl

247	5	3.1	64	1	US-08-344-695-14	Sequence 14, Appl
248	5	3.1	64	4	US-09-091-814-24	Sequence 24, Appl
249	5	3.1	64	4	US-09-091-814-25	Sequence 25, Appl
250	5	3.1	66	6	5252466-2	Patent No. 5252466
251	5	3.1	67	1	US-08-330-163-22	Sequence 22, Appl
252	5	3.1	67	1	US-08-482-111-22	Sequence 22, Appl
253	5	3.1	68	1	US-08-482-111-67	Sequence 67, Appl
254	5	3.1	69	4	US-08-866-545-11	Sequence 11, Appl
255	5	3.1	70	1	US-07-778-413E-21	Sequence 21, Appl
256	5	3.1	70	1	US-08-340-102-21	Sequence 21, Appl
257	5	3.1	70	1	US-08-330-163-6	Sequence 6, Appli
258	5	3.1	70	1	US-08-330-163-23	Sequence 23, Appl
259	5	3.1	70	1	US-08-330-163-24	Sequence 24, Appl
260	5	3.1	70	1	US-08-330-163-25	Sequence 25, Appl
261	5	3.1	70	1	US-08-330-163-26	Sequence 26, Appl
262	5	3.1	70	1	US-08-330-163-28	Sequence 28, Appl
263	5	3.1	70	1	US-08-321-071A-12	Sequence 12, Appl
264	5	3.1	70	1	US-08-259-550A-9	Sequence 9, Appli
265	5	3.1	70	1	US-08-482-111-6	Sequence 6, Appli
266	5	3.1	70	1	US-08-482-111-23	Sequence 23, Appl
267	5	3.1	70	1	US-08-482-111-24	Sequence 24, Appl
268	5	3.1	70	1	US-08-482-111-25	Sequence 25, Appl
269	5	3.1	70	1	US-08-482-111-26	Sequence 26, Appl
270	5	3.1	70	1	US-08-482-111-28	Sequence 28, Appl
271	5	3.1	70	2	US-08-436-420-30	Sequence 30, Appl
272	5	3.1	70	2	US-08-974-331-1	Sequence 1, Appli
273	5	3.1	70	2	US-08-803-734-1	Sequence 1, Appli
274	5	3.1	70	4	US-09-275-384B-3	Sequence 3, Appli
275	5	3.1	71	1	US-08-482-111-68	Sequence 68, Appl
276	5	3.1	71	6	5482923-2	Patent No. 5482923
277	5	3.1	76	2	US-08-436-420-40	Sequence 40, Appl
278	5	3.1	76	2	US-08-974-331-2	Sequence 2, Appli
279	5	3.1	77	2	US-08-102-385G-30	Sequence 30, Appl
280	5	3.1	77	4	US-09-134-001C-3500	Sequence 3500, Ap
281	5	3.1	77	4	US-09-152-060-67	Sequence 67, Appl
282	5	3.1	80	1	US-08-485-455D-59	Sequence 59, Appl
283	5	3.1	80	2	US-08-482-130C-59	Sequence 59, Appl
284	5	3.1	80	2	US-08-484-211C-59	Sequence 59, Appl
285	5	3.1	80	3	US-08-906-769-59	Sequence 59, Appl
286	5	3.1	80	3	US-08-906-616-59	Sequence 59, Appl
287	5	3.1	80	4	US-08-817-795-59	Sequence 59, Appl
288	5	3.1	80	4	US-08-485-443B-59	Sequence 59, Appl
289	5	3.1	80	4	US-08-639-075A-59	Sequence 59, Appl
290	5	3.1	80	4	US-09-012-431-59	Sequence 59, Appl
291	5	3.1	80	4	US-09-012-692-59	Sequence 59, Appl
292	5	3.1	80	4	US-08-906-613-59	Sequence 59, Appl
293	5	3.1	80	5	PCT-US95-14442A-59	Sequence 59, Appl
294	5	3.1	84	4	US-09-247-155-119	Sequence 119, App
295	5	3.1	85	2	US-08-222-617A-10	Sequence 10, Appl
296	5	3.1	87	2	US-08-460-934-7	Sequence 7, Appli
297	5	3.1	87	2	US-08-782-118-7	Sequence 7, Appli
298	5	3.1	87	3	US-09-195-868-22	Sequence 22, Appl
299	5	3.1	89	4	US-09-134-001C-5361	Sequence 5361, Ap
300	5	3.1	91	3	US-08-700-651-14	Sequence 14, Appl
301	5	3.1	91	3	US-08-928-361B-19	Sequence 19, Appl
302	5	3.1	91	4	US-08-905-223-409	Sequence 409, App
303	5	3.1	92	2	US-08-738-127-1	Sequence 1, Appli
304	5	3.1	92	4	US-09-247-155-120	Sequence 120, App
305	5	3.1	93	1	US-08-276-852-111	Sequence 111, App
306	5	3.1	93	1	US-08-899-575-111	Sequence 111, App
307	5	3.1	93	1	US-08-899-575-111	Sequence 111, App
308	5	3.1	93	2	US-08-467-822-36	Sequence 36, Appl
309	5	3.1	93	2	US-08-726-306A-142	Sequence 142, App
310	5	3.1	93	4	US-08-432-697-36	Sequence 36, Appl
311	5	3.1	93	4	US-08-466-248-36	Sequence 36, Appl
312	5	3.1	93	5	PCT-US95-08743-111	Sequence 111, App
313	5	3.1	97	1	US-08-202-389-18	Sequence 18, Appl
314	5	3.1	97	4	US-09-227-357-465	Sequence 465, App
315	5	3.1	98	1	US-08-118-270-330	Sequence 330, App
316	5	3.1	98	3	US-09-003-708A-6	Sequence 6, Appli
317	5	3.1	98	4	US-09-383-212-6	Sequence 6, Appli
318	5	3.1	98	5	PCT-US93-08528-330	Sequence 330, App
319	5	3.1	101	1	US-08-037-486-1	Sequence 1, Appli
320	5	3.1	101	1	US-08-476-008-17	Sequence 17, Appl
321	5	3.1	101	1	US-08-352-324A-8	Sequence 8, Appli
322	5	3.1	101	1	US-08-306-063-17	Sequence 17, Appl
323	5	3.1	101	1	US-08-833-485-17	Sequence 17, Appl
324	5	3.1	101	2	US-08-862-607-8	Sequence 8, Appli
325	5	3.1	101	3	US-09-203-235-8	Sequence 8, Appli
326	5	3.1	101	4	US-09-275-384B-2	Sequence 2, Appli
327	5	3.1	101	4	US-09-137-440-17	Sequence 17, Appl
328	5	3.1	101	4	US-09-347-801-2	Sequence 2, Appli
329	5	3.1	101	4	US-09-449-437A-8	Sequence 8, Appli
330	5	3.1	101	5	PCT-US91-06148A-17	Sequence 17, Appl
331	5	3.1	101	5	PCT-US95-16144-8	Sequence 8, Appli
332	5	3.1	104	6	5187075-5	Patent No. 5187075
333	5	3.1	105	2	US-08-918-727-6	Sequence 6, Appli
334	5	3.1	105	3	US-09-205-680A-6	Sequence 6, Appli
335	5	3.1	106	2	US-08-232-539D-54	Sequence 54, Appl
336	5	3.1	106	2	US-08-785-464-2	Sequence 2, Appli
337	5	3.1	107	1	US-08-487-748A-16	Sequence 16, Appl
338	5	3.1	107	3	US-08-398-633-16	Sequence 16, Appl
339	5	3.1	107	3	US-08-480-070C-16	Sequence 16, Appl
340	5	3.1	107	3	US-08-829-525-16	Sequence 16, Appl
341	5	3.1	107	4	US-08-609-583A-16	Sequence 16, Appl
342	5	3.1	107	4	US-09-240-274-162	Sequence 162, App
343	5	3.1	107	4	US-08-937-399-16	Sequence 16, Appl
344	5	3.1	107	4	US-09-310-367-16	Sequence 16, Appl
345	5	3.1	107	4	US-09-032-337-16	Sequence 16, Appl
346	5	3.1	108	1	US-08-276-852-110	Sequence 110, App
347	5	3.1	108	1	US-08-899-575-110	Sequence 110, App
348	5	3.1	108	1	US-08-899-575-110	Sequence 110, App
349	5	3.1	108	5	PCT-US95-08743-110	Sequence 110, App
350	5	3.1	109	2	US-08-675-508-6	Sequence 6, Appli
351	5	3.1	109	4	US-08-466-163B-1	Sequence 1, Appli
352	5	3.1	109	4	US-09-425-638A-20	Sequence 20, Appl
353	5	3.1	109	4	US-09-425-638A-21	Sequence 21, Appl
354	5	3.1	109	4	US-09-425-638A-39	Sequence 39, Appl
355	5	3.1	109	4	US-09-425-638A-40	Sequence 40, Appl
356	5	3.1	109	4	US-09-425-638A-41	Sequence 41, Appl
357	5	3.1	109	4	US-09-425-638A-42	Sequence 42, Appl
358	5	3.1	109	4	US-09-425-638A-43	Sequence 43, Appl
359	5	3.1	109	4	US-09-425-638A-44	Sequence 44, Appl
360	5	3.1	109	4	US-09-425-638A-45	Sequence 45, Appl
361	5	3.1	109	4	US-09-543-004-20	Sequence 20, Appl
362	5	3.1	109	4	US-09-543-004-21	Sequence 21, Appl
363	5	3.1	109	4	US-09-543-004-39	Sequence 39, Appl
364	5	3.1	109	4	US-09-543-004-40	Sequence 40, Appl
365	5	3.1	109	4	US-09-543-004-41	Sequence 41, Appl
366	5	3.1	109	4	US-09-543-004-42	Sequence 42, Appl
367	5	3.1	109	4	US-09-543-004-43	Sequence 43, Appl
368	5	3.1	109	4	US-09-543-004-44	Sequence 44, Appl
369	5	3.1	109	4	US-09-543-004-45	Sequence 45, Appl
370	5	3.1	109	4	US-09-187-859-11	Sequence 11, Appl
371	5	3.1	109	4	US-09-071-035-194	Sequence 194, App
372	5	3.1	110	1	US-08-487-748A-15	Sequence 15, Appl
373	5	3.1	110	3	US-08-398-633-15	Sequence 15, Appl
374	5	3.1	110	3	US-08-480-070C-15	Sequence 15, Appl
375	5	3.1	110	3	US-08-829-525-15	Sequence 15, Appl
376	5	3.1	110	4	US-08-609-583A-15	Sequence 15, Appl
377	5	3.1	110	4	US-08-937-399-15	Sequence 15, Appl
378	5	3.1	110	4	US-09-187-859-14	Sequence 14, Appl
379	5	3.1	110	4	US-09-310-367-15	Sequence 15, Appl
380	5	3.1	110	4	US-09-032-337-15	Sequence 15, Appl
381	5	3.1	111	4	US-09-134-001C-5398	Sequence 5398, Ap
382	5	3.1	112	4	US-09-395-689-6	Sequence 6, Appli
383	5	3.1	112	4	US-08-469-260A-68	Sequence 68, Appl
384	5	3.1	113	2	US-08-232-539D-56	Sequence 56, Appl
385	5	3.1	114	4	US-09-134-001C-3376	Sequence 3376, Ap
386	5	3.1	116	4	US-08-936-165A-304	Sequence 304, App
387	5	3.1	116	4	US-09-199-637A-409	Sequence 409, App
388	5	3.1	117	2	US-08-814-806-4	Sequence 4, Appli
389	5	3.1	117	4	US-09-344-587-12	Sequence 12, Appl
390	5	3.1	117	4	US-09-615-192A-400	Sequence 400, App
391	5	3.1	118	3	US-08-466-151-1	Sequence 1, Appli
392	5	3.1	119	2	US-08-464-025A-1	Sequence 1, Appli

393	5	3.1	121	3	US-08-747-221B-5	Sequence 5, Appli
394	5	3.1	121	4	US-09-005-051-5	Sequence 5, Appli
395	5	3.1	122	3	US-08-480-173A-56	Sequence 56, Appl
396	5	3.1	122	3	US-08-484-408A-56	Sequence 56, Appl
397	5	3.1	124	3	US-08-700-651-11	Sequence 11, Appl
398	5	3.1	124	3	US-08-928-361B-16	Sequence 16, Appl
399	5	3.1	124	4	US-09-134-001C-2929	Sequence 2929, Ap
400	5	3.1	125	1	US-08-462-949-24	Sequence 24, Appl
401	5	3.1	125	1	US-08-023-764B-24	Sequence 24, Appl
402	5	3.1	128	2	US-08-759-581B-21	Sequence 21, Appl
403	5	3.1	128	3	US-08-700-651-7	Sequence 7, Appli
404	5	3.1	128	3	US-08-928-361B-12	Sequence 12, Appl
405	5	3.1	128	4	US-09-304-711-21	Sequence 21, Appl
406	5	3.1	128	4	US-09-319-056B-5	Sequence 5, Appli
407	5	3.1	128	4	US-09-173-281-21	Sequence 21, Appl
408	5	3.1	129	4	US-08-942-686-4	Sequence 4, Appli
409	5	3.1	130	1	US-08-487-748A-13	Sequence 13, Appl
410	5	3.1	130	1	US-08-487-748A-14	Sequence 14, Appl
411	5	3.1	130	1	US-08-716-588-4	Sequence 4, Appli
412	5	3.1	130	1	US-08-716-588-5	Sequence 5, Appli
413	5	3.1	130	2	US-08-966-444-4	Sequence 4, Appli
414	5	3.1	130	2	US-08-966-444-5	Sequence 5, Appli
415	5	3.1	130	3	US-08-700-651-8	Sequence 8, Appli
416	5	3.1	130	3	US-08-700-651-9	Sequence 9, Appli
417	5	3.1	130	3	US-08-398-633-13	Sequence 13, Appl
418	5	3.1	130	3	US-08-398-633-14	Sequence 14, Appl
419	5	3.1	130	3	US-08-480-070C-13	Sequence 13, Appl
420	5	3.1	130	3	US-08-480-070C-14	Sequence 14, Appl
421	5	3.1	130	3	US-08-928-361B-13	Sequence 13, Appl
422	5	3.1	130	3	US-08-928-361B-14	Sequence 14, Appl
423	5	3.1	130	3	US-08-906-769-157	Sequence 157, App
424	5	3.1	130	3	US-08-829-525-13	Sequence 13, Appl
425	5	3.1	130	3	US-08-829-525-14	Sequence 14, Appl
426	5	3.1	130	3	US-08-906-616-157	Sequence 157, App
427	5	3.1	130	4	US-08-639-075A-157	Sequence 157, App
428	5	3.1	130	4	US-09-012-431-157	Sequence 157, App
429	5	3.1	130	4	US-08-609-583A-13	Sequence 13, Appl
430	5	3.1	130	4	US-08-609-583A-14	Sequence 14, Appl
431	5	3.1	130	4	US-09-012-692-157	Sequence 157, App
432	5	3.1	130	4	US-08-906-613-157	Sequence 157, App
433	5	3.1	130	4	US-08-937-399-13	Sequence 13, Appl
434	5	3.1	130	4	US-08-937-399-14	Sequence 14, Appl
435	5	3.1	130	4	US-08-339-214-28	Sequence 28, Appl
436	5	3.1	130	4	US-09-134-001C-3191	Sequence 3191, Ap
437	5	3.1	130	4	US-09-310-367-13	Sequence 13, Appl
438	5	3.1	130	4	US-09-310-367-14	Sequence 14, Appl
439	5	3.1	130	4	US-09-032-337-13	Sequence 13, Appl
440	5	3.1	130	4	US-09-032-337-14	Sequence 14, Appl
441	5	3.1	131	4	US-09-120-653D-3	Sequence 3, Appli
442	5	3.1	131	4	US-09-134-001C-4833	Sequence 4833, Ap
443	5	3.1	132	1	US-08-365-981-8	Sequence 8, Appli
444	5	3.1	132	4	US-08-826-134-8	Sequence 8, Appli
445	5	3.1	132	4	US-08-826-134-19	Sequence 19, Appl
446	5	3.1	133	1	US-08-196-218-32	Sequence 32, Appl
447	5	3.1	133	1	US-08-681-953-32	Sequence 32, Appl
448	5	3.1	133	4	US-09-134-001C-5352	Sequence 5352, Ap
449	5	3.1	134	4	US-09-465-558-48	Sequence 48, Appl
450	5	3.1	135	2	US-08-647-960-3	Sequence 3, Appli
451	5	3.1	136	4	US-08-339-214-83	Sequence 83, Appl
452	5	3.1	137	2	US-08-379-057-18	Sequence 18, Appl
453	5	3.1	137	4	US-08-339-214-2	Sequence 2, Appli
454	5	3.1	137	4	US-08-339-214-10	Sequence 10, Appl
455	5	3.1	137	4	US-08-339-214-18	Sequence 18, Appl
456	5	3.1	137	4	US-08-339-214-82	Sequence 82, Appl
457	5	3.1	138	3	US-08-700-651-10	Sequence 10, Appl
458	5	3.1	138	3	US-08-928-361B-15	Sequence 15, Appl
459	5	3.1	138	3	US-08-881-771A-2	Sequence 2, Appli
460	5	3.1	138	4	US-08-339-214-4	Sequence 4, Appli
461	5	3.1	138	4	US-08-339-214-12	Sequence 12, Appl
462	5	3.1	138	4	US-08-339-214-20	Sequence 20, Appl
463	5	3.1	138	4	US-08-339-214-81	Sequence 81, Appl
464	5	3.1	138	4	US-09-180-422B-12	Sequence 12, Appl
465	5	3.1	139	4	US-08-339-214-80	Sequence 80, Appl
466	5	3.1	140	4	US-08-569-147-76	Sequence 76, Appl
467	5	3.1	140	4	US-08-569-147-82	Sequence 82, Appl
468	5	3.1	141	4	US-09-286-529-22	Sequence 22, Appl
469	5	3.1	141	4	US-08-339-214-79	Sequence 79, Appl
470	5	3.1	142	4	US-09-370-838-213	Sequence 213, App
471	5	3.1	143	4	US-09-152-060-107	Sequence 107, App
472	5	3.1	145	3	US-08-630-172-5	Sequence 5, Appli
473	5	3.1	145	4	US-09-375-419-5	Sequence 5, Appli
474	5	3.1	146	4	US-08-751-359-15	Sequence 15, Appl
475	5	3.1	146	4	US-08-907-146-15	Sequence 15, Appl
476	5	3.1	147	3	US-08-904-871-7	Sequence 7, Appli
477	5	3.1	147	4	US-09-527-236A-19	Sequence 19, Appl
478	5	3.1	149	1	US-07-912-740D-47	Sequence 47, Appl
479	5	3.1	149	3	US-08-584-031-17	Sequence 17, Appl
480	5	3.1	150	1	US-07-912-740D-48	Sequence 48, Appl
481	5	3.1	150	3	US-08-928-361B-18	Sequence 18, Appl
482	5	3.1	150	4	US-09-134-001C-5062	Sequence 5062, Ap
483	5	3.1	150	4	US-09-373-953-4	Sequence 4, Appli
484	5	3.1	152	1	US-08-077-848A-4	Sequence 4, Appli
485	5	3.1	152	3	US-09-211-640-4	Sequence 4, Appli
486	5	3.1	152	4	US-09-378-536-4	Sequence 4, Appli
487	5	3.1	153	4	US-09-247-155-104	Sequence 104, App
488	5	3.1	154	1	US-08-363-010-1	Sequence 1, Appli
489	5	3.1	154	2	US-08-911-434A-4	Sequence 4, Appli
490	5	3.1	155	4	US-09-188-930-330	Sequence 330, App
491	5	3.1	155	4	US-08-935-433-4	Sequence 4, Appli
492	5	3.1	155	4	US-09-553-132-4	Sequence 4, Appli
493	5	3.1	155	4	US-09-589-287B-23	Sequence 23, Appl
494	5	3.1	156	1	US-08-074-121-2	Sequence 2, Appli
495	5	3.1	156	4	US-09-056-556-227	Sequence 227, App
496	5	3.1	156	4	US-09-134-001C-5462	Sequence 5462, Ap
497	5	3.1	156	4	US-09-615-192A-272	Sequence 272, App
498	5	3.1	156	4	US-09-072-596-222	Sequence 222, App
499	5	3.1	156	5	PCT-US94-06447-2	Sequence 2, Appli
500	5	3.1	157	4	US-09-364-230-26	Sequence 26, Appl
501	5	3.1	158	1	US-08-072-283B-13	Sequence 13, Appl
502	5	3.1	158	4	US-09-374-135-2	Sequence 2, Appli
503	5	3.1	158	4	US-08-858-207A-308	Sequence 308, App
504	5	3.1	162	3	US-08-700-651-13	Sequence 13, Appl
505	5	3.1	162	4	US-08-078-271B-6	Sequence 6, Appli
506	5	3.1	163	4	US-08-936-165A-420	Sequence 420, App
507	5	3.1	163	4	US-09-134-001C-4225	Sequence 4225, Ap
508	5	3.1	167	4	US-09-134-001C-4653	Sequence 4653, Ap
509	5	3.1	168	4	US-09-087-465-28	Sequence 28, Appl
510	5	3.1	168	4	US-08-975-762-37	Sequence 37, Appl
511	5	3.1	168	4	US-08-821-324-37	Sequence 37, Appl
512	5	3.1	168	4	US-09-295-028-37	Sequence 37, Appl
513	5	3.1	168	4	US-09-106-582-37	Sequence 37, Appl
514	5	3.1	168	4	US-09-376-113-5	Sequence 5, Appli
515	5	3.1	169	1	US-08-233-788A-53	Sequence 53, Appl
516	5	3.1	169	2	US-08-790-032-2	Sequence 2, Appli
517	5	3.1	169	3	US-09-069-619-2	Sequence 2, Appli
518	5	3.1	169	4	US-09-494-018-2	Sequence 2, Appli
519	5	3.1	170	3	US-09-130-663-2	Sequence 2, Appli
520	5	3.1	170	3	US-09-189-035-2	Sequence 2, Appli
521	5	3.1	170	3	US-09-069-896-1	Sequence 1, Appli
522	5	3.1	170	4	US-09-432-335-2	Sequence 2, Appli
523	5	3.1	170	4	US-09-382-086-2	Sequence 2, Appli
524	5	3.1	170	4	US-09-614-022-2	Sequence 2, Appli
525	5	3.1	170	4	US-09-471-468-1	Sequence 1, Appli
526	5	3.1	173	2	US-08-943-915-32	Sequence 32, Appl
527	5	3.1	173	4	US-09-724-864-57	Sequence 57, Appl
528	5	3.1	174	2	US-08-709-979A-6	Sequence 6, Appli
529	5	3.1	174	3	US-08-804-180C-4	Sequence 4, Appli
530	5	3.1	174	3	US-08-709-974A-9	Sequence 9, Appli
531	5	3.1	175	3	US-08-700-651-12	Sequence 12, Appl
532	5	3.1	175	3	US-08-928-361B-17	Sequence 17, Appl
533	5	3.1	175	4	US-09-376-113-7	Sequence 7, Appli
534	5	3.1	176	4	US-09-213-343-3	Sequence 3, Appli
535	5	3.1	178	1	US-08-271-354-7	Sequence 7, Appli
536	5	3.1	178	2	US-08-565-861-7	Sequence 7, Appli
537	5	3.1	178	3	US-09-122-443-15	Sequence 15, Appl
538	5	3.1	178	4	US-09-171-461-2	Sequence 2, Appli

539	5	3.1	178	4	US-08-339-214-84	Sequence 84, Appl
540	5	3.1	178	5	PCT-US94-07658-7	Sequence 7, Appli
541	5	3.1	179	3	US-08-649-100-9	Sequence 9, Appli
542	5	3.1	179	4	US-08-339-214-6	Sequence 6, Appli
543	5	3.1	179	4	US-08-339-214-14	Sequence 14, Appl
544	5	3.1	179	4	US-08-339-214-22	Sequence 22, Appl
545	5	3.1	181	1	US-08-471-058-18	Sequence 18, Appl
546	5	3.1	181	3	US-08-471-057-18	Sequence 18, Appl
547	5	3.1	182	2	US-08-417-495-25	Sequence 25, Appl
548	5	3.1	182	2	US-08-284-391B-25	Sequence 25, Appl
549	5	3.1	182	4	US-09-218-950-25	Sequence 25, Appl
550	5	3.1	182	5	PCT-US92-01785-25	Sequence 25, Appl
551	5	3.1	182	5	PCT-US95-00454-25	Sequence 25, Appl
552	5	3.1	183	4	US-09-311-352B-4	Sequence 4, Appli
553	5	3.1	183	4	US-09-373-953-2	Sequence 2, Appli
554	5	3.1	185	1	US-07-741-940-6	Sequence 6, Appli
555	5	3.1	185	1	US-08-289-548A-6	Sequence 6, Appli
556	5	3.1	185	1	US-08-452-654-6	Sequence 6, Appli
557	5	3.1	185	1	US-08-452-655B-6	Sequence 6, Appli
558	5	3.1	185	2	US-08-531-525-30	Sequence 30, Appl
559	5	3.1	185	2	US-08-718-270A-30	Sequence 30, Appl
560	5	3.1	185	2	US-08-865-336-3	Sequence 3, Appli
561	5	3.1	185	2	US-08-865-336-4	Sequence 4, Appli
562	5	3.1	185	3	US-08-450-582-6	Sequence 6, Appli
563	5	3.1	185	4	US-08-981-739-133	Sequence 133, App
564	5	3.1	185	4	US-08-981-739-136	Sequence 136, App
565	5	3.1	185	4	US-09-134-001C-3293	Sequence 3293, Ap
566	5	3.1	185	4	US-09-128-026-133	Sequence 133, App
567	5	3.1	185	4	US-09-128-026-136	Sequence 136, App
568	5	3.1	185	4	US-08-449-731-6	Sequence 6, Appli
569	5	3.1	186	1	US-08-698-551-6	Sequence 6, Appli
570	5	3.1	186	2	US-08-602-228-6	Sequence 6, Appli
571	5	3.1	186	2	US-08-649-341A-6	Sequence 6, Appli
572	5	3.1	186	2	US-08-494-440B-6	Sequence 6, Appli
573	5	3.1	186	2	US-08-533-901B-6	Sequence 6, Appli
574	5	3.1	186	2	US-08-839-032A-6	Sequence 6, Appli
575	5	3.1	186	2	US-08-839-031A-6	Sequence 6, Appli
576	5	3.1	186	4	US-09-185-258C-6	Sequence 6, Appli
577	5	3.1	186	5	PCT-US95-12724-6	Sequence 6, Appli
578	5	3.1	187	4	US-09-088-549-1	Sequence 1, Appli
579	5	3.1	190	1	US-08-044-621D-27	Sequence 27, Appl
580	5	3.1	190	1	US-08-393-985-25	Sequence 25, Appl
581	5	3.1	190	2	US-08-845-998-4	Sequence 4, Appli
582	5	3.1	190	2	US-08-845-998-6	Sequence 6, Appli
583	5	3.1	190	4	US-09-206-537-4	Sequence 4, Appli
584	5	3.1	190	4	US-09-206-537-6	Sequence 6, Appli
585	5	3.1	190	4	US-09-430-854-4	Sequence 4, Appli
586	5	3.1	190	4	US-09-430-854-6	Sequence 6, Appli
587	5	3.1	191	1	US-08-607-269-28	Sequence 28, Appl
588	5	3.1	191	5	PCT-US95-04600-28	Sequence 28, Appl
589	5	3.1	192	4	US-08-960-190A-38	Sequence 38, Appl
590	5	3.1	192	4	US-09-393-627B-22	Sequence 22, Appl
591	5	3.1	192	5	PCT-US94-05795-3	Sequence 3, Appli
592	5	3.1	195	4	US-09-475-316A-106	Sequence 106, App
593	5	3.1	195	4	US-09-285-677-2	Sequence 2, Appli
594	5	3.1	195	4	US-09-222-938A-58	Sequence 58, Appl
595	5	3.1	196	2	US-08-785-464-3	Sequence 3, Appli
596	5	3.1	196	2	US-08-943-208-2	Sequence 2, Appli
597	5	3.1	198	2	US-08-943-915-33	Sequence 33, Appl
598	5	3.1	199	2	US-08-831-575-9	Sequence 9, Appli
599	5	3.1	200	4	US-09-562-737-108	Sequence 108, App
600	5	3.1	200	5	PCT-US94-14277-4	Sequence 4, Appli
601	5	3.1	204	2	US-08-531-525-48	Sequence 48, Appl
602	5	3.1	204	2	US-08-718-270A-48	Sequence 48, Appl
603	5	3.1	204	6	5187075-6	Patent No. 5187075
604	5	3.1	205	1	US-08-277-231A-5	Sequence 5, Appli
605	5	3.1	205	2	US-08-473-750-8	Sequence 8, Appli
606	5	3.1	205	2	US-08-477-326-8	Sequence 8, Appli
607	5	3.1	207	2	US-08-808-550-29	Sequence 29, Appl
608	5	3.1	207	2	US-08-943-915-2	Sequence 2, Appli
609	5	3.1	207	2	US-08-943-915-5	Sequence 5, Appli
610	5	3.1	210	1	US-07-741-940-4	Sequence 4, Appli
611	5	3.1	210	1	US-08-320-559-4	Sequence 4, Appli
612	5	3.1	210	1	US-08-327-392-4	Sequence 4, Appli
613	5	3.1	210	1	US-08-289-548A-4	Sequence 4, Appli
614	5	3.1	210	1	US-08-452-654-4	Sequence 4, Appli
615	5	3.1	210	1	US-08-452-655B-4	Sequence 4, Appli
616	5	3.1	210	3	US-08-545-860D-4	Sequence 4, Appli
617	5	3.1	210	3	US-08-450-582-4	Sequence 4, Appli
618	5	3.1	210	4	US-08-449-731-4	Sequence 4, Appli
619	5	3.1	210	5	PCT-US94-04496-4	Sequence 4, Appli
620	5	3.1	211	2	US-08-808-550-38	Sequence 38, Appl
621	5	3.1	211	4	US-09-651-656-25	Sequence 25, Appl
622	5	3.1	211	4	US-09-650-855-25	Sequence 25, Appl
623	5	3.1	213	4	US-09-134-001C-5360	Sequence 5360, Ap
624	5	3.1	214	1	US-07-953-230A-11	Sequence 11, Appl
625	5	3.1	216	4	US-08-914-375C-4	Sequence 4, Appli
626	5	3.1	216	4	US-08-914-375C-17	Sequence 17, Appl
627	5	3.1	216	4	US-08-914-375C-19	Sequence 19, Appl
628	5	3.1	216	4	US-08-914-375C-20	Sequence 20, Appl
629	5	3.1	217	2	US-08-622-720A-24	Sequence 24, Appl
630	5	3.1	218	4	US-09-247-373B-6	Sequence 6, Appli
631	5	3.1	218	4	US-08-914-375C-6	Sequence 6, Appli
632	5	3.1	218	4	US-08-914-375C-11	Sequence 11, Appl
633	5	3.1	218	4	US-08-914-375C-12	Sequence 12, Appl
634	5	3.1	218	4	US-08-914-375C-13	Sequence 13, Appl
635	5	3.1	218	4	US-08-914-375C-14	Sequence 14, Appl
636	5	3.1	218	4	US-08-914-375C-15	Sequence 15, Appl
637	5	3.1	218	4	US-08-914-375C-16	Sequence 16, Appl
638	5	3.1	218	4	US-08-914-375C-30	Sequence 30, Appl
639	5	3.1	219	2	US-08-773-368-3	Sequence 3, Appli
640	5	3.1	219	3	US-08-924-747-6	Sequence 6, Appli
641	5	3.1	219	3	US-08-924-747-20	Sequence 20, Appl
642	5	3.1	219	3	US-09-199-887-3	Sequence 3, Appli
643	5	3.1	219	4	US-09-247-373B-20	Sequence 20, Appl
644	5	3.1	219	4	US-09-296-715-6	Sequence 6, Appli
645	5	3.1	219	4	US-09-296-715-20	Sequence 20, Appl
646	5	3.1	219	4	US-08-871-572B-12	Sequence 12, Appl
647	5	3.1	220	4	US-09-134-001C-5587	Sequence 5587, Ap
648	5	3.1	221	2	US-08-861-269-6	Sequence 6, Appli
649	5	3.1	221	2	US-09-134-596-6	Sequence 6, Appli
650	5	3.1	221	3	US-09-293-273-6	Sequence 6, Appli
651	5	3.1	223	3	US-08-228-208A-18	Sequence 18, Appl
652	5	3.1	223	4	US-09-254-733-7	Sequence 7, Appli
653	5	3.1	224	4	US-08-944-483-34	Sequence 34, Appl
654	5	3.1	224	4	US-08-858-207A-485	Sequence 485, App
655	5	3.1	224	4	US-09-134-001C-4010	Sequence 4010, Ap
656	5	3.1	224	5	PCT-US91-09055-4	Sequence 4, Appli
657	5	3.1	226	4	US-08-871-572B-10	Sequence 10, Appl
658	5	3.1	227	4	US-08-871-572B-14	Sequence 14, Appl
659	5	3.1	229	3	US-08-470-535-12	Sequence 12, Appl
660	5	3.1	229	4	US-09-399-913-70	Sequence 70, Appl
661	5	3.1	230	4	US-09-625-188-18	Sequence 18, Appl
662	5	3.1	230	4	US-09-485-737B-102	Sequence 102, App
663	5	3.1	231	2	US-08-336-198C-7	Sequence 7, Appli
664	5	3.1	231	3	US-08-908-332-6	Sequence 6, Appli
665	5	3.1	231	4	US-08-961-083-176	Sequence 176, App
666	5	3.1	232	2	US-08-956-047-36	Sequence 36, Appl
667	5	3.1	232	4	US-09-347-878-18	Sequence 18, Appl
668	5	3.1	233	2	US-08-928-692-31	Sequence 31, Appl
669	5	3.1	233	4	US-09-339-972-31	Sequence 31, Appl
670	5	3.1	233	4	US-09-485-737B-69	Sequence 69, Appl
671	5	3.1	234	1	US-08-505-058-2	Sequence 2, Appli
672	5	3.1	234	2	US-08-459-818-22	Sequence 22, Appl
673	5	3.1	234	2	US-08-889-666-22	Sequence 22, Appl
674	5	3.1	234	2	US-08-465-078-22	Sequence 22, Appl
675	5	3.1	234	2	US-08-725-776-22	Sequence 22, Appl
676	5	3.1	234	2	US-08-488-062-22	Sequence 22, Appl
677	5	3.1	235	4	US-08-862-540-9	Sequence 9, Appli
678	5	3.1	235	4	US-09-485-737B-93	Sequence 93, Appl
679	5	3.1	235	4	US-09-633-043-9	Sequence 9, Appli
680	5	3.1	236	1	US-08-315-695-22	Sequence 22, Appl
681	5	3.1	237	4	US-09-112-498A-7	Sequence 7, Appli
682	5	3.1	240	2	US-08-716-284-19	Sequence 19, Appl
683	5	3.1	240	4	US-09-485-737B-91	Sequence 91, Appl
684	5	3.1	241	4	US-08-944-483-59	Sequence 59, Appl

685	5	3.1	241	4	US-08-944-483-60	Sequence 60, Appl
686	5	3.1	242	3	US-08-908-332-2	Sequence 2, Appli
687	5	3.1	242	4	US-08-944-483-57	Sequence 57, Appl
688	5	3.1	242	4	US-08-944-483-58	Sequence 58, Appl
689	5	3.1	243	2	US-08-319-376-2	Sequence 2, Appli
690	5	3.1	244	1	US-08-696-827-1	Sequence 1, Appli
691	5	3.1	244	4	US-08-936-165A-379	Sequence 379, App
692	5	3.1	245	3	US-08-906-769-121	Sequence 121, App
693	5	3.1	245	3	US-08-906-616-121	Sequence 121, App
694	5	3.1	245	4	US-08-639-075A-121	Sequence 121, App
695	5	3.1	245	4	US-09-012-431-121	Sequence 121, App
696	5	3.1	245	4	US-09-012-692-121	Sequence 121, App
697	5	3.1	245	4	US-08-906-613-121	Sequence 121, App
698	5	3.1	245	4	US-09-079-970A-6	Sequence 6, Appli
699	5	3.1	246	4	US-09-216-295-20	Sequence 20, Appl
700	5	3.1	248	1	US-08-644-664B-27	Sequence 27, Appl
701	5	3.1	248	2	US-07-885-089B-6	Sequence 6, Appli
702	5	3.1	248	2	US-07-885-089B-10	Sequence 10, Appl
703	5	3.1	248	2	US-08-701-935-3	Sequence 3, Appli
704	5	3.1	248	2	US-08-761-277A-27	Sequence 27, Appl
705	5	3.1	248	3	US-09-134-591-3	Sequence 3, Appli
706	5	3.1	248	3	US-08-906-769-111	Sequence 111, App
707	5	3.1	248	3	US-08-906-616-111	Sequence 111, App
708	5	3.1	248	4	US-08-817-795-111	Sequence 111, App
709	5	3.1	248	4	US-08-639-075A-111	Sequence 111, App
710	5	3.1	248	4	US-09-012-431-111	Sequence 111, App
711	5	3.1	248	4	US-09-032-215-37	Sequence 37, Appl
712	5	3.1	248	4	US-09-012-692-111	Sequence 111, App
713	5	3.1	248	4	US-08-906-613-111	Sequence 111, App
714	5	3.1	248	5	PCT-US95-14442A-111	Sequence 111, App
715	5	3.1	249	3	US-08-700-651-15	Sequence 15, Appl
716	5	3.1	249	3	US-08-928-361B-20	Sequence 20, Appl
717	5	3.1	249	4	US-09-079-970A-5	Sequence 5, Appli
718	5	3.1	249	4	US-09-134-001C-5578	Sequence 5578, Ap
719	5	3.1	251	1	US-08-072-283B-12	Sequence 12, Appl
720	5	3.1	252	2	US-08-460-309-9	Sequence 9, Appli
721	5	3.1	252	2	US-08-125-077-9	Sequence 9, Appli
722	5	3.1	252	4	US-09-199-637A-176	Sequence 176, App
723	5	3.1	253	2	US-08-484-905-109	Sequence 109, App
724	5	3.1	253	3	US-08-481-985B-109	Sequence 109, App
725	5	3.1	253	4	US-08-370-476-109	Sequence 109, App
726	5	3.1	253	4	US-09-042-785A-4	Sequence 4, Appli
727	5	3.1	253	4	US-09-251-645-3	Sequence 3, Appli
728	5	3.1	254	2	US-08-701-935-6	Sequence 6, Appli
729	5	3.1	254	3	US-09-134-591-6	Sequence 6, Appli
730	5	3.1	254	4	US-08-944-483-50	Sequence 50, Appl
731	5	3.1	255	1	US-08-236-918A-8	Sequence 8, Appli
732	5	3.1	255	2	US-08-816-605-9	Sequence 9, Appli
733	5	3.1	255	2	US-07-690-192-4	Sequence 4, Appli
734	5	3.1	255	2	US-09-027-337-7	Sequence 7, Appli
735	5	3.1	255	4	US-09-006-353A-11	Sequence 11, Appl
736	5	3.1	255	4	US-09-007-097-2	Sequence 2, Appli
737	5	3.1	255	4	US-09-150-864A-8	Sequence 8, Appli
738	5	3.1	255	4	US-09-644-600-7	Sequence 7, Appli
739	5	3.1	255	4	US-09-573-986-11	Sequence 11, Appl
740	5	3.1	255	5	PCT-US96-03965-8	Sequence 8, Appli
741	5	3.1	256	2	US-08-286-819A-41	Sequence 41, Appl
742	5	3.1	256	2	US-08-484-905-114	Sequence 114, App
743	5	3.1	256	3	US-08-481-985B-114	Sequence 114, App
744	5	3.1	256	3	US-08-980-357-41	Sequence 41, Appl
745	5	3.1	256	4	US-08-370-476-114	Sequence 114, App
746	5	3.1	256	4	US-09-355-166-12	Sequence 12, Appl
747	5	3.1	257	3	US-08-728-603-19	Sequence 19, Appl
748	5	3.1	258	4	US-08-914-999-16	Sequence 16, Appl
749	5	3.1	258	4	US-09-134-001C-4647	Sequence 4647, Ap
750	5	3.1	259	2	US-08-484-905-117	Sequence 117, App
751	5	3.1	259	3	US-08-481-985B-117	Sequence 117, App
752	5	3.1	259	4	US-08-370-476-117	Sequence 117, App
753	5	3.1	260	4	US-08-743-168B-2	Sequence 2, Appli
754	5	3.1	260	5	PCT-US96-10435-2	Sequence 2, Appli
755	5	3.1	261	4	US-09-468-738A-29	Sequence 29, Appl
756	5	3.1	261	4	US-09-940-019-29	Sequence 29, Appl
757	5	3.1	262	4	US-08-154-364-41	Sequence 41, Appl
758	5	3.1	262	4	US-09-372-422A-32	Sequence 32, Appl
759	5	3.1	263	1	US-08-393-985-23	Sequence 23, Appl
760	5	3.1	263	2	US-08-685-992-13	Sequence 13, Appl
761	5	3.1	263	2	US-09-144-925-13	Sequence 13, Appl
762	5	3.1	263	4	US-08-818-112-91	Sequence 91, Appl
763	5	3.1	263	4	US-08-818-111-92	Sequence 92, Appl
764	5	3.1	263	4	US-09-056-556-91	Sequence 91, Appl
765	5	3.1	263	4	US-09-072-596-92	Sequence 92, Appl
766	5	3.1	265	4	US-09-227-357-153	Sequence 153, App
767	5	3.1	266	2	US-08-773-368-4	Sequence 4, Appli
768	5	3.1	266	3	US-09-199-887-4	Sequence 4, Appli
769	5	3.1	267	2	US-09-016-366A-23	Sequence 23, Appl
770	5	3.1	267	2	US-08-978-404B-18	Sequence 18, Appl
771	5	3.1	267	4	US-09-118-442-17	Sequence 17, Appl
772	5	3.1	267	4	US-09-677-064-17	Sequence 17, Appl
773	5	3.1	267	4	US-09-485-737B-2	Sequence 2, Appli
774	5	3.1	268	2	US-08-424-641B-8	Sequence 8, Appli
775	5	3.1	268	2	US-08-820-980-8	Sequence 8, Appli
776	5	3.1	268	2	US-08-826-439-8	Sequence 8, Appli
777	5	3.1	269	2	US-08-424-641B-3	Sequence 3, Appli
778	5	3.1	269	2	US-08-820-980-3	Sequence 3, Appli
779	5	3.1	269	2	US-08-826-439-3	Sequence 3, Appli
780	5	3.1	269	4	US-08-913-159-4	Sequence 4, Appli
781	5	3.1	270	1	US-08-414-926A-27	Sequence 27, Appl
782	5	3.1	270	2	US-08-926-922-27	Sequence 27, Appl
783	5	3.1	270	3	US-09-253-682-27	Sequence 27, Appl
784	5	3.1	270	4	US-09-527-657-27	Sequence 27, Appl
785	5	3.1	271	5	PCT-US92-01196-3	Sequence 3, Appli
786	5	3.1	272	1	US-08-698-551-8	Sequence 8, Appli
787	5	3.1	272	2	US-08-602-228-8	Sequence 8, Appli
788	5	3.1	272	2	US-08-649-341A-8	Sequence 8, Appli
789	5	3.1	272	2	US-08-494-440B-8	Sequence 8, Appli
790	5	3.1	272	2	US-08-533-901B-8	Sequence 8, Appli
791	5	3.1	272	2	US-08-839-032A-8	Sequence 8, Appli
792	5	3.1	272	2	US-08-839-031A-8	Sequence 8, Appli
793	5	3.1	272	4	US-09-185-258C-8	Sequence 3, Appli
794	5	3.1	272	4	US-08-456-640-3	Sequence 3, Appli
795	5	3.1	272	4	US-08-460-221-3	Sequence 4, Appli
796	5	3.1	272	5	PCT-US92-01196-4	Sequence 8, Appli
797	5	3.1	272	5	PCT-US95-12724-8	Sequence 5, Appli
798	5	3.1	273	2	US-08-701-935-5	Sequence 19, Appl
799	5	3.1	273	2	US-09-016-366A-19	Sequence 14, Appl
800	5	3.1	273	2	US-08-978-404B-14	Sequence 5, Appli
801	5	3.1	273	3	US-09-134-591-5	Sequence 21, Appl
802	5	3.1	274	2	US-09-016-366A-21	Sequence 16, Appl
803	5	3.1	274	2	US-08-978-404B-16	Sequence 105, App
804	5	3.1	274	2	US-08-484-905-105	Sequence 106, App
805	5	3.1	274	2	US-08-484-905-106	Sequence 107, App
806	5	3.1	274	2	US-08-484-905-107	Sequence 108, App
807	5	3.1	274	2	US-08-484-905-108	Sequence 105, App
808	5	3.1	274	3	US-08-481-985B-105	Sequence 106, App
809	5	3.1	274	3	US-08-481-985B-106	Sequence 107, App
810	5	3.1	274	3	US-08-481-985B-107	Sequence 108, App
811	5	3.1	274	3	US-08-481-985B-108	Sequence 105, App
812	5	3.1	274	4	US-08-370-476-105	Sequence 106, App
813	5	3.1	274	4	US-08-370-476-106	Sequence 107, App
814	5	3.1	274	4	US-08-370-476-107	Sequence 108, App
815	5	3.1	274	4	US-08-370-476-108	Sequence 3, Appli
816	5	3.1	274	5	PCT-US96-10435-3	Sequence 5, Appli
817	5	3.1	277	1	US-08-762-129-5	Sequence 16, Appl
818	5	3.1	277	2	US-08-567-375-16	Sequence 9, Appli
819	5	3.1	278	1	US-08-188-582-9	Sequence 9, Appli
820	5	3.1	278	1	US-08-646-715-9	Sequence 2, Appli
821	5	3.1	278	2	US-08-722-626B-2	Sequence 2, Appli
822	5	3.1	278	3	US-09-260-283-2	Sequence 16, Appl
823	5	3.1	278	4	US-08-339-214-16	Sequence 26, Appl
824	5	3.1	278	4	US-08-339-214-26	Sequence 701, App
825	5	3.1	278	4	US-09-149-476-701	Sequence 24, Appl
826	5	3.1	279	4	US-08-339-214-24	Sequence 32, Appl
827	5	3.1	279	4	US-08-339-214-32	Sequence 5, Appli
828	5	3.1	279	5	PCT-US95-00362-5	Sequence 2, Appli
829	5	3.1	281	2	US-08-810-453-2	Sequence 2, Appli
830	5	3.1	281	3	US-08-815-190A-2	

831	5	3.1	281	4	US-09-290-640-25	Sequence 25, Appl
832	5	3.1	281	4	US-09-479-524-3	Sequence 3, Appli
833	5	3.1	281	4	US-08-339-214-8	Sequence 8, Appli
834	5	3.1	281	4	US-08-339-214-30	Sequence 30, Appl
835	5	3.1	281	4	US-09-134-001C-3849	Sequence 3849, Ap
836	5	3.1	281	4	US-09-589-287B-6	Sequence 6, Appli
837	5	3.1	281	4	US-09-157-864-10	Sequence 10, Appl
838	5	3.1	281	4	US-09-006-755B-1	Sequence 1, Appli
839	5	3.1	281	5	PCT-US95-00362-2	Sequence 2, Appli
840	5	3.1	283	4	US-08-992-035A-3	Sequence 3, Appli
841	5	3.1	284	4	US-08-564-164A-2	Sequence 2, Appli
842	5	3.1	285	1	US-08-393-985-2	Sequence 2, Appli
843	5	3.1	285	3	US-09-027-137-3	Sequence 2, Appli
844	5	3.1	285	4	US-08-992-035A-1	Sequence 3, Appli
845	5	3.1	285	4	US-09-344-441-3	Sequence 1, Appli
846	5	3.1	286	2	US-08-809-267-3	Sequence 3, Appli
847	5	3.1	286	5	PCT-US95-13662A-3	Sequence 3, Appli
848	5	3.1	287	2	US-08-741-437-5	Sequence 3, Appli
849	5	3.1	287	2	US-09-134-593-5	Sequence 5, Appli
850	5	3.1	287	3	US-08-815-190A-16	Sequence 5, Appli
851	5	3.1	288	1	US-08-337-602-4	Sequence 16, Appl
852	5	3.1	288	3	US-08-558-135-4	Sequence 4, Appli
853	5	3.1	288	4	US-08-819-286-3	Sequence 4, Appli
854	5	3.1	289	4	US-09-215-252-10	Sequence 3, Appli
855	5	3.1	289	4	US-09-134-001C-4929	Sequence 10, Appl
856	5	3.1	289	4	US-09-222-938A-19	Sequence 4929, Ap
857	5	3.1	290	4	US-09-386-653A-7	Sequence 19, Appl
858	5	3.1	292	2	US-08-403-852D-25	Sequence 7, Appli
859	5	3.1	292	3	US-08-510-646B-26	Sequence 25, Appl
860	5	3.1	292	4	US-09-231-818-25	Sequence 26, Appl
861	5	3.1	292	4	US-08-765-907A-17	Sequence 25, Appl
862	5	3.1	294	2	US-08-923-856-3	Sequence 17, Appl
863	5	3.1	294	3	US-09-216-294-3	Sequence 3, Appli
864	5	3.1	294	6	5260223-1	Sequence 3, Appli
865	5	3.1	295	4	US-08-893-654B-6	Patent No. 5260223
866	5	3.1	296	4	US-09-134-001C-2976	Sequence 6, Appli
867	5	3.1	300	2	US-08-946-528-7	Sequence 2976, Ap
868	5	3.1	300	4	US-09-277-078-2	Sequence 7, Appli
869	5	3.1	301	2	US-08-785-464-1	Sequence 2, Appli
870	5	3.1	304	4	US-09-088-651-2	Sequence 1, Appli
871	5	3.1	304	4	US-09-105-390-40	Sequence 2, Appli
872	5	3.1	305	1	US-08-266-451B-26	Sequence 40, Appl
873	5	3.1	305	2	US-08-748-725-26	Sequence 26, Appl
874	5	3.1	306	4	US-09-451-900-1	Sequence 26, Appl
875	5	3.1	306	4	US-09-921-318-1	Sequence 1, Appli
876	5	3.1	307	2	US-08-807-861A-56	Sequence 1, Appli
877	5	3.1	307	3	US-09-210-681-56	Sequence 56, Appl
878	5	3.1	307	3	US-08-946-719A-56	Sequence 56, Appl
879	5	3.1	308	4	US-08-887-534A-26	Sequence 56, Appl
880	5	3.1	309	1	US-08-078-683A-5	Sequence 26, Appl
881	5	3.1	309	2	US-08-997-080-118	Sequence 5, Appli
882	5	3.1	309	2	US-08-997-362-118	Sequence 118, App
883	5	3.1	309	4	US-09-095-855-118	Sequence 118, App
884	5	3.1	309	4	US-08-818-112-74	Sequence 118, App
885	5	3.1	309	4	US-09-324-542-118	Sequence 74, Appl
886	5	3.1	309	4	US-08-818-111-75	Sequence 118, App
887	5	3.1	309	4	US-09-056-556-74	Sequence 75, Appl
888	5	3.1	309	4	US-09-205-426-118	Sequence 74, Appl
889	5	3.1	309	4	US-09-072-596-75	Sequence 118, App
890	5	3.1	310	1	US-08-078-683A-3	Sequence 75, Appl
891	5	3.1	310	2	US-08-488-199-4	Sequence 3, Appli
892	5	3.1	310	3	US-08-651-136C-22	Sequence 4, Appli
893	5	3.1	310	4	US-09-229-911A-22	Sequence 22, Appl
894	5	3.1	311	1	US-08-078-683A-2	Sequence 22, Appl
895	5	3.1	311	1	US-08-472-217-2	Sequence 2, Appli
896	5	3.1	311	2	US-08-488-199-6	Sequence 2, Appli
897	5	3.1	311	3	US-08-760-534A-2	Sequence 6, Appli
898	5	3.1	311	4	US-09-105-390-44	Sequence 2, Appli
899	5	3.1	311	4	US-09-134-001C-4235	Sequence 44, Appl
900	5	3.1	312	3	US-08-910-501-2	Sequence 4235, Ap
901	5	3.1	312	3	US-08-910-501-4	Sequence 2, Appli
902	5	3.1	312	4	US-09-398-550-2	Sequence 4, Appli
903	5	3.1	312	4	US-09-398-550-4	Sequence 2, Appli
904	5	3.1	313	1	US-08-078-683A-4	Sequence 4, Appli
905	5	3.1	313	3	US-08-686-528A-3	Sequence 3, Appli
906	5	3.1	313	4	US-09-475-316A-62	Sequence 62, Appl
907	5	3.1	313	4	US-09-456-287-3	Sequence 3, Appli
908	5	3.1	315	4	US-09-434-774-16	Sequence 16, Appl
909	5	3.1	315	4	US-09-386-653A-9	Sequence 9, Appli
910	5	3.1	317	2	US-09-066-075-2	Sequence 2, Appli
911	5	3.1	317	2	US-08-518-615A-2	Sequence 2, Appli
912	5	3.1	317	3	US-08-951-889-2	Sequence 2, Appli
913	5	3.1	317	4	US-09-472-857-2	Sequence 2, Appli
914	5	3.1	318	1	US-08-220-958-4	Sequence 4, Appli
915	5	3.1	318	4	US-09-060-756-727	Sequence 727, App
916	5	3.1	320	3	US-08-793-666-1	Sequence 1, Appli
917	5	3.1	320	3	US-08-793-666-10	Sequence 10, Appl
918	5	3.1	320	4	US-09-134-001C-4802	Sequence 4802, Ap
919	5	3.1	321	2	US-08-813-940-6	Sequence 6, Appli
920	5	3.1	321	4	US-08-748-506-10	Sequence 10, Appl
921	5	3.1	321	4	US-08-748-506-18	Sequence 18, Appl
922	5	3.1	322	2	US-08-428-414A-2	Sequence 2, Appli
923	5	3.1	324	2	US-08-813-940-2	Sequence 2, Appli
924	5	3.1	324	4	US-09-222-938A-31	Sequence 31, Appl
925	5	3.1	325	4	US-09-041-886-33	Sequence 33, Appl
926	5	3.1	325	4	US-09-088-435-1	Sequence 1, Appli
927	5	3.1	326	1	US-07-603-133B-26	Sequence 26, Appl
928	5	3.1	326	4	US-09-134-001C-4721	Sequence 4721, Ap
929	5	3.1	326	6	5223424-2	Patent No. 5223424
930	5	3.1	327	1	US-08-240-049B-13	Sequence 13, Appl
931	5	3.1	327	1	US-08-240-049B-14	Sequence 14, Appl
932	5	3.1	327	1	US-08-259-148A-15	Sequence 15, Appl
933	5	3.1	327	1	US-08-259-148A-16	Sequence 16, Appl
934	5	3.1	327	1	US-08-484-054-15	Sequence 15, Appl
935	5	3.1	327	1	US-08-484-054-16	Sequence 16, Appl
936	5	3.1	327	2	US-07-876-941A-15	Sequence 15, Appl
937	5	3.1	327	2	US-07-876-941A-16	Sequence 16, Appl
938	5	3.1	327	2	US-08-997-080-162	Sequence 162, App
939	5	3.1	327	2	US-08-997-362-162	Sequence 162, App
940	5	3.1	327	4	US-09-095-855-162	Sequence 162, App
941	5	3.1	327	4	US-08-542-634-17	Sequence 17, Appl
942	5	3.1	327	4	US-08-542-634-18	Sequence 18, Appl
943	5	3.1	327	4	US-08-477-292-17	Sequence 17, Appl
944	5	3.1	327	4	US-08-477-292-18	Sequence 18, Appl
945	5	3.1	327	4	US-09-324-542-162	Sequence 162, App
946	5	3.1	327	4	US-09-134-001C-3535	Sequence 3535, Ap
947	5	3.1	327	4	US-09-205-426-162	Sequence 162, App
948	5	3.1	327	4	US-07-870-985A-15	Sequence 15, Appl
949	5	3.1	327	4	US-07-870-985A-16	Sequence 16, Appl
950	5	3.1	327	5	PCT-US95-13703-17	Sequence 17, Appl
951	5	3.1	327	5	PCT-US95-13703-18	Sequence 18, Appl
952	5	3.1	328	2	US-08-977-767-1	Sequence 1, Appli
953	5	3.1	328	4	US-09-615-192A-274	Sequence 274, App
954	5	3.1	329	4	US-09-134-001C-3120	Sequence 3120, Ap
955	5	3.1	330	4	US-09-145-391-2	Sequence 2, Appli
956	5	3.1	331	4	US-09-105-390-56	Sequence 56, Appl
957	5	3.1	332	2	US-08-469-537A-75	Sequence 75, Appl
958	5	3.1	332	4	US-09-134-001C-4015	Sequence 4015, Ap
959	5	3.1	333	1	US-08-440-856A-4	Sequence 4, Appli
960	5	3.1	333	1	US-08-436-463-6	Sequence 6, Appli
961	5	3.1	333	1	US-08-024-253-6	Sequence 6, Appli
962	5	3.1	333	4	US-09-562-737-2	Sequence 2, Appli
963	5	3.1	334	2	US-08-665-647-3	Sequence 3, Appli
964	5	3.1	335	4	US-09-222-938A-4	Sequence 4, Appli
965	5	3.1	335	4	US-09-641-259B-25	Sequence 25, Appl
966	5	3.1	336	4	US-09-105-390-60	Sequence 60, Appl
967	5	3.1	337	1	US-08-118-270-33	Sequence 33, Appl
968	5	3.1	337	1	US-08-440-856A-3	Sequence 3, Appli
969	5	3.1	337	3	US-08-686-528A-2	Sequence 2, Appli
970	5	3.1	337	4	US-09-456-287-2	Sequence 2, Appli
971	5	3.1	337	5	PCT-US93-08528-33	Sequence 33, Appl
972	5	3.1	339	4	US-09-232-191-2	Sequence 2, Appli
973	5	3.1	339	4	US-09-232-200-2	Sequence 2, Appli
974	5	3.1	339	4	US-09-232-197-2	Sequence 2, Appli
975	5	3.1	339	4	US-09-232-201-2	Sequence 2, Appli
976	5	3.1	340	2	US-08-102-385G-18	Sequence 18, Appl

977	5	3.1	340	4	US-08-975-762-54	Sequence 54, Appl
978	5	3.1	340	4	US-09-295-028-54	Sequence 54, Appl
979	5	3.1	340	4	US-09-106-582-54	Sequence 54, Appl
980	5	3.1	341	1	US-07-748-783-4	Sequence 4, Appli
981	5	3.1	341	1	US-08-166-818-4	Sequence 4, Appli
982	5	3.1	341	4	US-09-134-001C-3783	Sequence 3783, Ap
983	5	3.1	342	2	US-08-899-011-2	Sequence 2, Appli
984	5	3.1	342	4	US-09-373-958-2	Sequence 2, Appli
985	5	3.1	342	4	US-09-381-810A-1	Sequence 1, Appli
986	5	3.1	342	4	US-09-129-033-2	Sequence 2, Appli
987	5	3.1	342	4	US-09-134-001C-5198	Sequence 5198, Ap
988	5	3.1	343	4	US-08-853-948B-3	Sequence 3, Appli
989	5	3.1	344	2	US-07-857-224B-92	Sequence 92, Appl
990	5	3.1	344	2	US-07-857-224B-93	Sequence 93, Appl
991	5	3.1	344	4	US-08-818-112-69	Sequence 69, Appl
992	5	3.1	344	4	US-08-818-111-70	Sequence 70, Appl
993	5	3.1	344	4	US-09-056-556-69	Sequence 69, Appl
994	5	3.1	344	4	US-09-134-001C-4304	Sequence 4304, Ap
995	5	3.1	344	4	US-09-072-596-70	Sequence 70, Appl
996	5	3.1	345	3	US-08-858-003-34	Sequence 34, Appl
997	5	3.1	345	3	US-09-078-166-34	Sequence 34, Appl
998	5	3.1	345	4	US-08-997-467-34	Sequence 34, Appl
999	5	3.1	346	4	US-09-199-637A-313	Sequence 313, App
1000	5	3.1	348	3	US-08-875-540-13	Sequence 13, Appl

ALIGNMENTS

```
RESULT 1
US-08-409-731A-11
; Sequence 11, Application US/08409731A
; Patent No. 5658758
;  GENERAL INFORMATION:
;    APPLICANT:  Ni, Jian
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Gentz, Reiner
;    APPLICANT:  Rosen, Craig A.
;    TITLE OF INVENTION:  CYTOSTATIN I
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES, INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/409,731A
;      FILING DATE:  24-MAR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Robert H
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PF175
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  301-309-8504
;      TELEFAX:  301-309-8512
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  132 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-409-731A-11

  Query Match             4.3%;  Score 7;  DB 1;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 VTLQRGS 37
           |||||||
Db      85 VTLQRGS 91


RESULT 2
US-08-470-298B-11
; Sequence 11, Application US/08470298B
; Patent No. 5844081
;  GENERAL INFORMATION:
;    APPLICANT:  NI, JIAN
;    APPLICANT:  GENTZ, REINER
;    APPLICANT:  YU, GUO-LIANG
;    APPLICANT:  ROSEN, CRAIG A.
;    TITLE OF INVENTION:  CYTOSTATIN I
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES, INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MD
;      COUNTRY:  US
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,298B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROOKES, ALLAN A.
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF175D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  301-309-8504
;      TELEFAX:  301-309-8512
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  132 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  MYELIN P2 (FIGURE 2)
US-08-470-298B-11

  Query Match             4.3%;  Score 7;  DB 2;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 VTLQRGS 37
           |||||||
Db      85 VTLQRGS 91


RESULT 3
US-09-023-073A-11
; Sequence 11, Application US/09023073A
; Patent No. 5977309
;  GENERAL INFORMATION:
;    APPLICANT:  Ni, Jian
;    APPLICANT:  Gentz, Reiner
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Rosen, Craig A
;    TITLE OF INVENTION:  Cytostatin I
;    NUMBER OF SEQUENCES:  11
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES, INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,073A
;      FILING DATE:  13-FEB-1998
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wales, Michele M.
;      REGISTRATION NUMBER:  P-43,975
;      REFERENCE/DOCKET NUMBER:  PF175D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  301-610-5772
;      TELEFAX:  301-309-8439
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  132 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-023-073A-11

  Query Match             4.3%;  Score 7;  DB 2;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 VTLQRGS 37
           |||||||
Db      85 VTLQRGS 91


RESULT 4
US-09-361-737-11
; Sequence 11, Application US/09361737
; Patent No. 6287812
;  GENERAL INFORMATION:
;    APPLICANT:  Ni, Jian
;    APPLICANT:  Gentz, Reiner
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Rosen, Craig A
;    TITLE OF INVENTION:  Cytostatin I
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES, INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/361,737
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/023,073
;      FILING DATE:  13-FEB-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Wales, Michele M.
;      REGISTRATION NUMBER:  P-43,975
;      REFERENCE/DOCKET NUMBER:  PF175D2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  301-610-5772
;      TELEFAX:  301-309-8439
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  132 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-361-737-11

  Query Match             4.3%;  Score 7;  DB 4;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 VTLQRGS 37
           |||||||
Db      85 VTLQRGS 91


RESULT 5
US-08-132-990A-8
; Sequence 8, Application US/08132990A
; Patent No. 5834589
;  GENERAL INFORMATION:
;    APPLICANT:  MERUELO, DANIEL
;    APPLICANT:  YOSHIMOTO, TAKAYUKI
;    TITLE OF INVENTION:  Human Retrovirus Receptor and DNA Coding Therefor
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent In Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/132,990A
;      FILING DATE:  07-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/084,729
;      FILING DATE:  29-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/05569
;      FILING DATE:  11-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/899,075
;      FILING DATE:  11-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/806,178
;      FILING DATE:  13-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/627,950
;      FILING DATE:  14-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  8105-004-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66441 PENNIE
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  629 amino acids
```

```
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-132-990A-8

  Query Match              4.3%;  Score 7;  DB 2;  Length 629;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 SLIAVLI 105
           |||||||
Db     501 SLIAVLI 507


RESULT 6
PCT-US92-09382-8
; Sequence 8, Application PC/TUS9209382
;  GENERAL INFORMATION:
;    APPLICANT:  MERUELO, DANIEL
;    APPLICANT:  YOSHIMOTO, TAKAYUKI
;    TITLE OF INVENTION:  Human Retrovirus Receptor and DNA Coding
;    TITLE OF INVENTION:  Therefor
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Browdy and Neimark
;      STREET:  419 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/09382
;      FILING DATE:  19921213
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Livnat, Shmuel
;      REGISTRATION NUMBER:  33,949
;      REFERENCE/DOCKET NUMBER:  MERUELO=1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202 628-5197
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  629 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-09382-8

  Query Match              4.3%;  Score 7;  DB 5;  Length 629;
  Best Local Similarity   100.0%;  Pred. No. 73;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 SLIAVLI 105
           |||||||
Db     501 SLIAVLI 507


RESULT 7
US-09-134-001C-4014
; Sequence 4014, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4014
;    LENGTH: 92
;    TYPE: PRT
;    ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4014

  Query Match              3.7%;  Score 6;  DB 4;  Length 92;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 NLVVDL 13
            ||||||
Db       6 NLVVDL 11


RESULT 8
US-08-326-117B-13
; Sequence 13, Application US/08326117B
; Patent No. 5693491
;  GENERAL INFORMATION:
;    APPLICANT:  BULLA, LEE A.
;    APPLICANT:  JI, TAE
;    TITLE OF INVENTION:  RECEPTOR FOR A BACILLUS THURINGIENSIS
;    TITLE OF INVENTION:  TOXIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Ave. N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/326,117B
;      FILING DATE:  19-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MILLMAN, ROBERT A
;      REGISTRATION NUMBER:  36,217
;      REFERENCE/DOCKET NUMBER:  7112-0037.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  124 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-326-117B-13

  Query Match              3.7%;  Score 6;  DB 1;  Length 124;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLV 94
           ||||||
Db      15 SAGGLV 20


RESULT 9
US-08-982-129-13
```

```
; Sequence 13, Application US/08982129
; Patent No. 6007981
; GENERAL INFORMATION:
;   APPLICANT: BULLA, LEE A.
;   APPLICANT: JI, TAE
;   TITLE OF INVENTION: RECEPTOR FOR A BACILLUS THURINGIENSIS
;   TITLE OF INVENTION: TOXIN
;   NUMBER OF SEQUENCES: 26
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: MORRISON & FOERSTER
;     STREET: 2000 Pennsylvania Ave. N.W.
;     CITY: Washington
;     STATE: DC
;     COUNTRY: USA
;     ZIP: 20006-1812
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/982,129
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/326,117
;     FILING DATE: 19-OCT-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: MILLMAN, ROBERT A
;     REGISTRATION NUMBER: 36,217
;     REFERENCE/DOCKET NUMBER: 7112-0037.00
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (202) 887-1500
;     TELEFAX: (202) 887-0763
;     TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 124 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
US-08-982-129-13

  Query Match             3.7%;  Score 6;  DB 3;  Length 124;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLV 94
           ||||||
Db      15 SAGGLV 20


RESULT 10
US-08-983-607-51
; Sequence 51, Application US/08983607
; Patent No. 6140470
; GENERAL INFORMATION:
;   APPLICANT: Alan Garen
;   APPLICANT: Xiaohong Cai
;   TITLE OF INVENTION: Human Anti-Tumor Monoclonal Anti-
;   TITLE OF INVENTION: bodies
;   NUMBER OF SEQUENCES: 51
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Department of Molecular Biophysics
;     ADDRESSEE: and Biochemistry, Yale University
;     STREET: 266 Whitney Avenue
;     CITY: New Haven
;     STATE: Connecticut
;     COUNTRY: United States of America
;     ZIP: 06520-8114
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: 3.5" 1.44 Mb diskette
;     COMPUTER: IBM PC
;     OPERATING SYSTEM: MS DOS
;     SOFTWARE: Word Processing
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/983,607
;     FILING DATE: April 27, 1998
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PCT/IB96/01032
;     FILING DATE: June 28, 1996
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Mary M. Krinsky
;     REGISTRATION NUMBER: 32423
;     REFERENCE/DOCKET NUMBER: OCR-679
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 203-773-9544
;     TELEFAX: 203-773-1183
; INFORMATION FOR SEQ ID NO: 51:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 124 residues
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE:
;     DESCRIPTION: polypeptide
;   ORIGINAL SOURCE:
;     ORGANISM: Homo sapiens (melanoma patient
;     ORGANISM: immunized with autologous tumor cells)
;     INDIVIDUAL ISOLATE: peripheral blood lympho-
;     INDIVIDUAL ISOLATE: cytes
;   IMMEDIATE SOURCE:
;     LIBRARY: VH antibodies obtained from fUSE5
;     LIBRARY: fusion phage construct
;     CLONE: E-13
;   FEATURE:
;     NAME/KEY: heavy chain
US-08-983-607-51

  Query Match             3.7%;  Score 6;  DB 4;  Length 124;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLV 94
           ||||||
Db       7 SAGGLV 12


RESULT 11
US-09-134-001C-4330
; Sequence 4330, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE: 1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4330
;   LENGTH: 146
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4330

  Query Match             3.7%;  Score 6;  DB 4;  Length 146;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      45 NFSGTV 50
           ||||||
Db      77 NFSGTV 82


RESULT 12
US-08-985-526-36
; Sequence 36, Application US/08985526
; Patent No. 6080728
;  GENERAL INFORMATION:
;    APPLICANT:  Mixson, James A
;    TITLE OF INVENTION:  CARRIER:DNA COMPLEXES CONTAINING DNA
;    TITLE OF INVENTION:  ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
;    TITLE OF INVENTION:  THERAPY
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Connolly, Bove, Lodge, & Hutz
;      STREET:  1220 Market Street, P.O. Box 2207
;      CITY:  Wilmington
;      STATE:  Delaware
;      COUNTRY:  U.S.A.
;      ZIP:  19899
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/985,526
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/608,845
;      FILING DATE:  16-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMorrow Jr., Robert G
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (302) 658-9141
;      TELEFAX:  (302) 658-5613
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  185 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-985-526-36

  Query Match             3.7%;  Score 6;  DB 3;  Length 185;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     151 RGSVPI 156
           ||||||
Db      66 RGSVPI 71


RESULT 13
US-09-561-500-13
; Sequence 13, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE:  4001.002500
;  CURRENT APPLICATION NUMBER: US/09/561,500
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-500-13

  Query Match             3.7%;  Score 6;  DB 4;  Length 191;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     151 RGSVPI 156
           ||||||
Db      73 RGSVPI 78


RESULT 14
US-09-561-108-13
; Sequence 13, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VE
;  FILE REFERENCE:  4001.002584
;  CURRENT APPLICATION NUMBER: US/09/561,108
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-108-13

  Query Match             3.7%;  Score 6;  DB 4;  Length 191;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     151 RGSVPI 156
           ||||||
Db      73 RGSVPI 78


RESULT 15
US-09-561-526-13
; Sequence 13, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
;  APPLICANT:  Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION:  ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
;  FILE REFERENCE:  4001.002586
;  CURRENT APPLICATION NUMBER: US/09/561,526
;  CURRENT FILING DATE:  2000-04-28
;  PRIOR APPLICATION NUMBER:  60/131,432
;  PRIOR FILING DATE:  1999-04-28
;  NUMBER OF SEQ ID NOS:  44
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-561-526-13

  Query Match             3.7%;  Score 6;  DB 4;  Length 191;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    151 RGSVPI 156
          ||||||
Db     73 RGSVPI 78


RESULT 16
US-08-159-784-2
; Sequence 2, Application US/08159784
; Patent No. 5643783
;  GENERAL INFORMATION:
;    APPLICANT:  Bjorn R. Olsen
;    TITLE OF INVENTION:  NOVEL COLLAGEN AND USES THEREOF
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      COMPUTER:  IBM PS/2 Model 50Z or 55SX
;      OPERATING SYSTEM:  MS-DOS (Version 5.0)
;      SOFTWARE:  WordPerfect (Version 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/159,784
;      FILING DATE:  December 1, 1993
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  John F. Freeman
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  00246/170001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  195
;      TYPE:  amino acid
;      STRANDEDNESS:  N/A
;      TOPOLOGY:  N/A
US-08-159-784-2

  Query Match             3.7%;  Score 6;  DB 1;  Length 195;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    151 RGSVPI 156
          ||||||
Db     77 RGSVPI 82


RESULT 17
US-08-248-839C-94
; Sequence 94, Application US/08248839C
; Patent No. 5843702
;  GENERAL INFORMATION:
;    APPLICANT:  McConnell, David
;    APPLICANT:  Devine, Kevin
;    APPLICANT:  O'Kane, Charles
;    TITLE OF INVENTION:  A Gene Expression System
;    NUMBER OF SEQUENCES:  185
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5843702o No. 5843702disk of No. 5843702th America, Inc.
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/248,839C
;      FILING DATE:  25-MAY-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gregg, Valeta A.
;      REGISTRATION NUMBER:  35,127
;      REFERENCE/DOCKET NUMBER:  3614.214-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  94:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  213 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Protein
US-08-248-839C-94

  Query Match             3.7%;  Score 6;  DB 2;  Length 213;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     27 ITDYVT 32
          ||||||
Db    148 ITDYVT 153


RESULT 18
US-07-885-089B-4
; Sequence 4, Application US/07885089B
; Patent No. 5830995
;  GENERAL INFORMATION:
;    APPLICANT:  Shoyab, Mohammed
;    APPLICANT:  McDonald, Vicki L.
;    APPLICANT:  Bradley, James G.
;    APPLICANT:  Plowman, Gregory D.
;    TITLE OF INVENTION:  AMPHIREGULINS:  A FAMILY OF
;    TITLE OF INVENTION:  HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,089B
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  5624-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  243 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-885-089B-4

  Query Match              3.7%;  Score 6;  DB 2;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLV 94
           ||||||
Db      45 SAGGLV 50


RESULT 19
US-07-885-089B-9
; Sequence 9, Application US/07885089B
; Patent No. 5830995
;  GENERAL INFORMATION:
;    APPLICANT:  Shoyab, Mohammed
;    APPLICANT:  McDonald, Vicki L.
;    APPLICANT:  Bradley, James G.
;    APPLICANT:  Plowman, Gregory D.
;    TITLE OF INVENTION:  AMPHIREGULINS:  A FAMILY OF
;    TITLE OF INVENTION:  HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,089B
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  5624-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-07-885-089B-9

  Query Match              3.7%;  Score 6;  DB 2;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLV 94
           ||||||
Db      45 SAGGLV 50


RESULT 20
US-08-944-483-70
; Sequence 70, Application US/08944483
; Patent No. 6232456
;  GENERAL INFORMATION:
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STEWART, KENT D.
;    APPLICANT:  STROUPE, STEVEN D.
;    TITLE OF INVENTION:  NOVEL SERINE PROTEASE REAGENTS
;    TITLE OF INVENTION:  AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
;    TITLE OF INVENTION:  OF THE PROSTATE
;    NUMBER OF SEQUENCES:  76
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,483
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6183.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  70:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6232456e
US-08-944-483-70

  Query Match              3.7%;  Score 6;  DB 4;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      97 AGSLIA 102
           ||||||
Db      26 AGSLIA 31


RESULT 21
5202428-2
;Patent No. 5202428
;    APPLICANT: SCHUBERT, DAVID
;    TITLE OF INVENTION: DNA ENCODING NEUROTROPHIC GROWTH FACTOR
;    NUMBER OF SEQUENCES: 10
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/590,359
;      FILING DATE: 27-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 541,276
;      FILING DATE: 20-JUN-1990
;SEQ ID NO:2:
;      LENGTH: 243
5202428-2
```

```
  Query Match             3.7%;  Score 6;  DB 6;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLV 94
           ||||||
Db      45 SAGGLV 50


RESULT 22
US-08-956-267A-2
; Sequence 2, Application US/08956267A
; Patent No. 5945328
;  GENERAL INFORMATION:
;    APPLICANT:  WOLDIKE, Helle Fabricius
;    APPLICANT:  KJELDSEN, Thomas Borglum
;    TITLE OF INVENTION:  A Process For Producing Trypsin
;    TITLE OF INVENTION:  (Trypsinogen)
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5945328o No. 5945328disk of No. 5945328th America, Inc.
;      STREET:  405 Lexington Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10174
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/956,267A
;      FILING DATE:  22-OCT-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rozek, Carol
;      REGISTRATION NUMBER:  36,993
;      REFERENCE/DOCKET NUMBER:  4500.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  247 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-956-267A-2

  Query Match             3.7%;  Score 6;  DB 2;  Length 247;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 SGSSYP 58
           ||||||
Db     151 SGSSYP 156


RESULT 23
US-08-989-386-5
; Sequence 5, Application US/08989386
; Patent No. 5989860
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Guegler, Karl J.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN ISOMERASE HOMOLOGS
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/989,386
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0443 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  249 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  KIDNNOT09
;      CLONE:  1419071
US-08-989-386-5

  Query Match             3.7%;  Score 6;  DB 2;  Length 249;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     137 DVSARD 142
           ||||||
Db     125 DVSARD 130


RESULT 24
US-09-238-303-10
; Sequence 10, Application US/09238303B
; Patent No. 6284253
; GENERAL INFORMATION:
;  APPLICANT: Barr, Margaret C.
;  TITLE OF INVENTION: No. 6284253el Feline Immunodeficiency Virus Nucleotide Sequenc
;  FILE REFERENCE: 18617.0059
;  CURRENT APPLICATION NUMBER: US/09/238,303B
;  CURRENT FILING DATE: 1999-01-28
;  EARLIER APPLICATION NUMBER: US 60/072,927
;  EARLIER FILING DATE: 1998-01-29
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 10
;   LENGTH: 252
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: protein encoded by the vif gene of a recombinant viral
;   OTHER INFORMATION: clone constructed from the genomic DNA of a Pallas's cat felin
;   OTHER INFORMATION: immunodeficiency virus
US-09-238-303-10

  Query Match             3.7%;  Score 6;  DB 4;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      92 GLVIKA 97
           ||||||
Db     154 GLVIKA 159


RESULT 25
US-08-491-944-2
; Sequence 2, Application US/08491944
; Patent No. 6361982
;   GENERAL INFORMATION:
;        APPLICANT: Walters, Nicola Jane
;                   Barton, Barry
;                   Earl, Alison J.
;        TITLE OF INVENTION: No. 6361982el Compounds
;        NUMBER OF SEQUENCES: 2
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: SmithKline Beecham p.l.c.
;             STREET: 709 Swedeland Road
;             CITY: King of Prussia
;             STATE: PA
;             COUNTRY: USA
;             ZIP: 19406-0939
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/491,944
;             FILING DATE: 03-Aug-1995
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Gimmi, Edward R
;             REGISTRATION NUMBER: 38,891
;             REFERENCE/DOCKET NUMBER: P31457-1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 610-270-4478
;             TELEFAX: 610-270-5090
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 262 amino acids
;             TYPE: amino acid
;             TOPOLOGY: unknown
;        MOLECULE TYPE: protein
;        HYPOTHETICAL: YES
;        ORIGINAL SOURCE:
;             ORGANISM: Streptomyces clavuligerus
;             STRAIN: S. clavuligerus 27064
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-491-944-2

  Query Match             3.7%;  Score 6;  DB 4;  Length 262;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     152 GSVPIP 157
           ||||||
Db      20 GSVPIP 25


RESULT 26
US-09-025-769B-178
; Sequence 178, Application US/09025769B
; Patent No. 6300064
;  GENERAL INFORMATION:
;    APPLICANT:  Knappik, Achim
;    APPLICANT:  Pack, Peter
;    APPLICANT:  Ilag, Vic
;    APPLICANT:  Ge, Liming
;    APPLICANT:  Moroney, Simon
;    APPLICANT:  Plueckthun, Andreas
;    TITLE OF INVENTION:  Protein/(Poly)peptide libraries
;    NUMBER OF SEQUENCES:  373
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  James F. Haley, Jr., Esq. c/o Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10021
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/025,769B
;      FILING DATE:  18-FEB-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 95 11 3021.0
;      FILING DATE:  18-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  James F. Haley, Jr., Esq.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  MORPHO/5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)596-9000
;      TELEFAX:  (212)596-9090
;  INFORMATION FOR SEQ ID NO:  178:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  281 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-025-769B-178

  Query Match             3.7%;  Score 6;  DB 4;  Length 281;
  Best Local Similarity   100.0%;  Pred. No. 3.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      87 VSSAGG 92
           ||||||
Db     143 VSSAGG 148


RESULT 27
US-09-512-342-14
; Sequence 14, Application US/09512342
; Patent No. 6388068
; GENERAL INFORMATION:
;  APPLICANT:  SATOH, SHINOBU
;  APPLICANT:  MASUDA, SUSUMU
;  TITLE OF INVENTION:  METHOD FOR PRODUCING FOREIGN POLYPEPTIDE IN PLANT
;  TITLE OF INVENTION:  INTERCELLULAR FLUID
;  FILE REFERENCE:  081356/0142
;  CURRENT APPLICATION NUMBER: US/09/512,342
;  CURRENT FILING DATE:  2000-02-24
;  NUMBER OF SEQ ID NOS:  38
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 293
;   TYPE: PRT
;   ORGANISM: Cucumis sativus
US-09-512-342-14

  Query Match             3.7%;  Score 6;  DB 4;  Length 293;
  Best Local Similarity   100.0%;  Pred. No. 3.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      29 DYVTLQ 34
           ||||||
Db     144 DYVTLQ 149


RESULT 28
```

```
US-08-118-270-57
; Sequence 57, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
;   APPLICANT: Murphy, Randall B.
;   APPLICANT: Schuster, David I.
;   TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
;   TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
;   NUMBER OF SEQUENCES: 348
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: BROWDY AND NEIMARK
;     STREET: 419 Seventh Street, N.W., Suite 300
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20004
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/118,270
;     FILING DATE: 09-SEP-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/943,236
;     FILING DATE: 10-SEP-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Townsend, Kevin G.
;     REGISTRATION NUMBER: 34,033
;     REFERENCE/DOCKET NUMBER: MURPHY=2A
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-628-5197
;     TELEFAX: 202-737-3528
;     TELEX: 248633
; INFORMATION FOR SEQ ID NO: 57:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 297 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-118-270-57

  Query Match              3.7%;  Score 6;  DB 1;  Length 297;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 SGSSYP 58
           ||||||
Db     150 SGSSYP 155


RESULT 29
US-08-118-270-58
; Sequence 58, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
;   APPLICANT: Murphy, Randall B.
;   APPLICANT: Schuster, David I.
;   TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
;   TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
;   NUMBER OF SEQUENCES: 348
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: BROWDY AND NEIMARK
;     STREET: 419 Seventh Street, N.W., Suite 300
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20004
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/118,270
;     FILING DATE: 09-SEP-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/943,236
;     FILING DATE: 10-SEP-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Townsend, Kevin G.
;     REGISTRATION NUMBER: 34,033
;     REFERENCE/DOCKET NUMBER: MURPHY=2A
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-628-5197
;     TELEFAX: 202-737-3528
;     TELEX: 248633
; INFORMATION FOR SEQ ID NO: 58:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 297 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: peptide
US-08-118-270-58

  Query Match              3.7%;  Score 6;  DB 1;  Length 297;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 SGSSYP 58
           ||||||
Db     149 SGSSYP 154


RESULT 30
PCT-US93-08528-57
; Sequence 57, Application PC/TUS9308528
; GENERAL INFORMATION:
;   APPLICANT: New York University
;   TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
;   TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
;   NUMBER OF SEQUENCES: 348
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: BROWDY AND NEIMARK
;     STREET: 419 Seventh Street, N.W., Suite 300
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20004
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US93/08528
;     FILING DATE: 09-SEP-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/943,236
;     FILING DATE: 10-SEP-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Townsend, Kevin G.
;     REGISTRATION NUMBER: 34,033
;     REFERENCE/DOCKET NUMBER: MURPHY=2 PCT
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-628-5197
;     TELEFAX: 202-737-3528
;     TELEX: 248633
; INFORMATION FOR SEQ ID NO: 57:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 297 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US93-08528-57

  Query Match              3.7%;  Score 6;  DB 5;  Length 297;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 SGSSYP 58
           ||||||
Db     150 SGSSYP 155


RESULT 31
PCT-US93-08528-58
; Sequence 58, Application PC/TUS9308528
;  GENERAL INFORMATION:
;    APPLICANT:  New York University
;    TITLE OF INVENTION:  POLYPEPTIDES OF G-COUPLED PROTEIN
;    TITLE OF INVENTION:  RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
;    NUMBER OF SEQUENCES:  348
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08528
;      FILING DATE:  09-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/943,236
;      FILING DATE:  10-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Townsend, Kevin G.
;      REGISTRATION NUMBER:  34,033
;      REFERENCE/DOCKET NUMBER:  MURPHY=2 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  58:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  297 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US93-08528-58

  Query Match              3.7%;  Score 6;  DB 5;  Length 297;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      53 SGSSYP 58
           ||||||
Db     149 SGSSYP 154


RESULT 32
US-09-134-001C-3238
; Sequence 3238, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3238
;   LENGTH: 307
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3238

  Query Match              3.7%;  Score 6;  DB 4;  Length 307;
  Best Local Similarity   100.0%;  Pred. No. 3.7e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     112 YNSDDF 117
           ||||||
Db     192 YNSDDF 197


RESULT 33
US-09-424-349A-8
; Sequence 8, Application US/09424349A
; Patent No. 6387668
;   GENERAL INFORMATION:
;        APPLICANT: Lutje Spelberg, Jeffrey Harald
;                   Rink, Rick
;                   Kellogg, Richard Morrison
;                   Janssen, Dick Barend
;        TITLE OF INVENTION: Enantioselective epoxide hydrolases and genes
;                            encoding these.
;        NUMBER OF SEQUENCES: 9
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Vereenigde
;             STREET: Nieuwe Parklaan 97
;             CITY: The Hague
;             STATE: Zuid-Holland
;             COUNTRY: The Netherlands
;             ZIP: 2587 BN
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: ASCII
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/424,349A
;             FILING DATE: 23-Feb-2000
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: EP 97201515.0
;             FILING DATE: 21-MAY-1997
;             APPLICATION NUMBER: PCT/NL98/00290
;             FILING DATE: 20-MAY-1998
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 329 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: unknown
;             TOPOLOGY: unknown
;        MOLECULE TYPE: peptide
;        HYPOTHETICAL: NO
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-424-349A-8

  Query Match              3.7%;  Score 6;  DB 4;  Length 329;
  Best Local Similarity   100.0%;  Pred. No. 3.9e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 VLSNFS 47
           ||||||
Db     127 VLSNFS 132
```

```
RESULT 34
US-08-164-292B-13
; Sequence 13, Application US/08164292B
; Patent No. 5820868
;  GENERAL INFORMATION:
;    APPLICANT:  MITTAL, SURESH K.
;    APPLICANT:  GRAHAM, FRANK L.
;    APPLICANT:  PREVEC, LUDVIK
;    APPLICANT:  BABIUK, LORNE A.
;    TITLE OF INVENTION:  RECOMBINANT PROTEIN PRODUCTION IN BOVINE
;    TITLE OF INVENTION:  ADENOVIRUS EXPRESSION VECTOR SYSTEM
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  345 California Street
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94104-2675
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/164,292B
;      FILING DATE:  09-DEC-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GRACEY, NANCY J.
;      REGISTRATION NUMBER:  28,216
;      REFERENCE/DOCKET NUMBER:  29310-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 677-7000
;      TELEFAX:  (415) 677-7522
;      TELEX:  34-0154
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  347 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-164-292B-13

  Query Match             3.7%;  Score 6;  DB 2;  Length 347;
  Best Local Similarity   100.0%;  Pred. No. 4.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      45 NFSGTV 50
           ||||||
Db      96 NFSGTV 101


RESULT 35
US-08-845-623-13
; Sequence 13, Application US/08845623A
; Patent No. 6001591
; GENERAL INFORMATION:
;  APPLICANT: BABIUK, LORNE A.
;  APPLICANT: TIKOO, SURESH K.
;  APPLICANT: REDDY, POLICE S.
;  TITLE OF INVENTION: BOVINE ADENOVIRUS 3 GENOME
;  FILE REFERENCE: 293102002120
;  CURRENT APPLICATION NUMBER: US/08/845,623A
;  CURRENT FILING DATE: 1997-04-25
;  EARLIER APPLICATION NUMBER: 08/164,294
;  EARLIER FILING DATE: 1993-12-09
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 347
;   TYPE: PRT
;   ORGANISM: Human adenovirus type 5
US-08-845-623-13

  Query Match             3.7%;  Score 6;  DB 3;  Length 347;
  Best Local Similarity   100.0%;  Pred. No. 4.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      45 NFSGTV 50
           ||||||
Db      96 NFSGTV 101


RESULT 36
US-08-815-927-13
; Sequence 13, Application US/08815927
; Patent No. 6086890
; GENERAL INFORMATION:
;  APPLICANT: MITTAL, SURESH K.
;  APPLICANT: GRAHAM, FRANK L.
;  APPLICANT: PREVIC, LUDVIK
;  APPLICANT: BABIUK, LORNE A.
;  TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS EXPRESSION
;  TITLE OF INVENTION: VECTOR SYSTEM
;  FILE REFERENCE: 293102002101
;  CURRENT APPLICATION NUMBER: US/08/815,927
;  CURRENT FILING DATE: 1997-03-13
;  EARLIER APPLICATION NUMBER: 08/164,294
;  EARLIER FILING DATE: 1993-12-09
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 347
;   TYPE: PRT
;   ORGANISM: Human adenovirus type 5
US-08-815-927-13

  Query Match             3.7%;  Score 6;  DB 3;  Length 347;
  Best Local Similarity   100.0%;  Pred. No. 4.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      45 NFSGTV 50
           ||||||
Db      96 NFSGTV 101


RESULT 37
US-09-103-330-13
; Sequence 13, Application US/09103330A
; Patent No. 6319716
; GENERAL INFORMATION:
;  APPLICANT: TIKOO, SURESH K.
;  APPLICANT: BABIUK, LORNE A.
;  APPLICANT: REDDY, POLICE S.
;  TITLE OF INVENTION: ISOLATION OF MUTANTS IN THE E3 REGION OF THE
;  TITLE OF INVENTION: BOVINE ADENOVIRUS GENOME AND THEIR USE IN VACCINES
;  FILE REFERENCE: 293102002121
;  CURRENT APPLICATION NUMBER: US/09/103,330A
;  CURRENT FILING DATE: 1998-06-23
;  EARLIER APPLICATION NUMBER: 08/880,234
;  EARLIER FILING DATE: 1997-06-23
;  EARLIER APPLICATION NUMBER: 08/164,292
;  EARLIER FILING DATE: 1993-12-09
;  NUMBER OF SEQ ID NOS: 40
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 347
;   TYPE: PRT
;   ORGANISM: Human adenovirus type 5
US-09-103-330-13

  Query Match             3.7%;  Score 6;  DB 4;  Length 347;
```

```
  Best Local Similarity   100.0%;  Pred. No. 4.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      45 NFSGTV 50
           ||||||
Db      96 NFSGTV 101


RESULT 38
US-09-435-242-13
; Sequence 13, Application US/09435242
; Patent No. 6379944
; GENERAL INFORMATION:
;  APPLICANT: MITTAL, SURESH K.
;  APPLICANT: GRAHAM, FRANK L.
;  APPLICANT: PREVIC, LUDVIK
;  APPLICANT: BABIUK, LORNE A.
;  TITLE OF INVENTION: MAMMALIAN CELL LINES EXPRESSING BOVINE ADENOVIRUS FUNCTIONS
;  FILE REFERENCE: 293102002102
;  CURRENT APPLICATION NUMBER: US/09/435,242
;  CURRENT FILING DATE: 1999-11-05
;  EARLIER APPLICATION NUMBER: 08/815,927
;  EARLIER FILING DATE: 1997-03-13
;  EARLIER APPLICATION NUMBER: 08/164,294
;  EARLIER FILING DATE: 1993-12-09
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 347
;   TYPE: PRT
;   ORGANISM: Human adenovirus type 5
US-09-435-242-13

  Query Match             3.7%;  Score 6;  DB 4;  Length 347;
  Best Local Similarity   100.0%;  Pred. No. 4.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      45 NFSGTV 50
           ||||||
Db      96 NFSGTV 101


RESULT 39
US-09-615-192A-338
; Sequence 338, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
;  APPLICANT:  Bloksberg, Leonard N.
;  APPLICANT:  Havukkala, Ilkka
;  TITLE OF INVENTION:  Materials and Methods for the
;  TITLE OF INVENTION:  Modification of Plant Lignin Content
;  FILE REFERENCE:  11000.1003c4U
;  CURRENT APPLICATION NUMBER: US/09/615,192A
;  CURRENT FILING DATE:  2000-07-12
;  PRIOR APPLICATION NUMBER:  US 08/975,316
;  PRIOR FILING DATE:  1997-11-21
;  PRIOR APPLICATION NUMBER:  US 08/713,000
;  PRIOR FILING DATE:  1996-09-11
;  PRIOR APPLICATION NUMBER:  US 09/169,789
;  PRIOR FILING DATE:  1998-10-09
;  NUMBER OF SEQ ID NOS:  405
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 338
;   LENGTH: 358
;   TYPE: PRT
;   ORGANISM: Pinus radiata
US-09-615-192A-338

  Query Match             3.7%;  Score 6;  DB 4;  Length 358;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      54 GSSYPF 59
              ||||||
Db      45 GSSYPF 50


RESULT 40
US-09-134-001C-4365
; Sequence 4365, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4365
;   LENGTH: 366
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4365

  Query Match             3.7%;  Score 6;  DB 4;  Length 366;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      83 YLTPVS 88
           ||||||
Db      39 YLTPVS 44


RESULT 41
US-09-134-001C-4259
; Sequence 4259, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4259
;   LENGTH: 376
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4259

  Query Match             3.7%;  Score 6;  DB 4;  Length 376;
  Best Local Similarity   100.0%;  Pred. No. 4.4e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      21 NDYPET 26
           ||||||
Db     247 NDYPET 252


RESULT 42
US-08-689-421-23
; Sequence 23, Application US/08689421
; Patent No. 6008029
;  GENERAL INFORMATION:
;    APPLICANT:  Yaver, Debbie S.
;    APPLICANT:  Brown, Kimberley M.
```

```
;    APPLICANT:  Kauppinen, Sakari
;    APPLICANT:  Halkier, Torben P
;    TITLE OF INVENTION:  PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS
;    TITLE OF INVENTION:  ENCODING SAME
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6008029o No. 6008029disk of No. 6008029th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/689,421
;      FILING DATE:  9-AUG-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  4554.204-WO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  387 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-689-421-23

  Query Match             3.7%;  Score 6;  DB 3;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VVNVGQ 7
           ||||||
Db     361 VVNVGQ 366


RESULT 43
US-09-389-528-23
; Sequence 23, Application US/09389528
; Patent No. 6207430
;  GENERAL INFORMATION:
;    APPLICANT:  Yaver, Debbie S.
;    APPLICANT:  Brown, Kimberley M.
;    APPLICANT:  Kauppinen, Sakari
;    APPLICANT:  Halkier, Torben P
;    TITLE OF INVENTION:  PURIFIED COPRINUS LACCASES AND NUCLEIC ACIDS
;    TITLE OF INVENTION:  ENCODING SAME
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6207430o No. 6207430disk of No. 6207430th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/389,528
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/689,421
;      FILING DATE:  9-AUG-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  4554.204-WO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  387 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-389-528-23

  Query Match             3.7%;  Score 6;  DB 4;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VVNVGQ 7
           ||||||
Db     361 VVNVGQ 366


RESULT 44
US-09-181-827A-23
; Sequence 23, Application US/09181827A
; Patent No. 6242232
; GENERAL INFORMATION:
;  APPLICANT:  Yaver, Debbie S.
;  APPLICANT:  Brown, Kimberly M.
;  APPLICANT:  Kauppinen, Sakari
;  APPLICANT:  Halkier, Torben P.
;  TITLE OF INVENTION:  Purified Coprinus Laccases And Nucleic
;  TITLE OF INVENTION:  Acids Encoding Same
;  FILE REFERENCE:  4554.200-US
;  CURRENT APPLICATION NUMBER: US/09/181,827A
;  CURRENT FILING DATE:  1998-10-28
;  PRIOR APPLICATION NUMBER:  60/002,800
;  PRIOR FILING DATE:  1995-08-25
;  NUMBER OF SEQ ID NOS:  36
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 23
;   LENGTH: 387
;   TYPE: PRT
;   ORGANISM: Coprinus cinereus
US-09-181-827A-23

  Query Match             3.7%;  Score 6;  DB 4;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 4.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 VVNVGQ 7
           ||||||
Db     361 VVNVGQ 366


RESULT 45
US-09-131-028A-4
; Sequence 4, Application US/09131028A
; Patent No. 6287866
; GENERAL INFORMATION:
;  APPLICANT: Abbott Laboratories
;  APPLICANT:  Mukerji, Pradip
;  APPLICANT:  Lemmel, Steven A.
;  APPLICANT:  Leonard, Amanda Eun-Yeong
;  APPLICANT:  Chaudhary, Sunita
;  TITLE OF INVENTION: BETA-CASEIN EXPRESSING CONSTRUCTS
;  FILE REFERENCE: 6004.US.P1
;  CURRENT APPLICATION NUMBER: US/09/131,028A
```

```
;  CURRENT FILING DATE:  1998-08-07
;  PRIOR APPLICATION NUMBER: US 08/064,440
;  PRIOR FILING DATE: 1993-05-21
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
;' SEQ ID NO 4
;   LENGTH: 391
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-131-028A-4

  Query Match              3.7%;  Score 6;  DB 4;  Length 391;
  Best Local Similarity   100.0%;  Pred. No. 4.6e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TPVSSA 90
           ||||||
Db     344 TPVSSA 349


Search completed: November 28, 2002, 19:06:41
Job time : 25 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 28, 2002, 19:05:41 ; Search time 13 Seconds
(without alignments)
197.215 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 161
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 102317 seqs, 15924203 residues

Word size : 0

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	161	100.0	279	10	US-09-900-575-29	Sequence 29, Appl
2	92	57.1	279	10	US-09-900-575-34	Sequence 34, Appl
3	92	57.1	279	10	US-09-900-575-36	Sequence 36, Appl
4	92	57.1	279	10	US-09-900-575-44	Sequence 44, Appl
5	92	57.1	300	10	US-09-912-020-367	Sequence 367, App
6	90	55.9	279	10	US-09-900-575-27	Sequence 27, Appl
7	90	55.9	279	10	US-09-900-575-28	Sequence 28, Appl
8	90	55.9	279	10	US-09-900-575-32	Sequence 32, Appl
9	90	55.9	279	10	US-09-900-575-39	Sequence 39, Appl
10	90	55.9	279	10	US-09-900-575-43	Sequence 43, Appl
11	90	55.9	279	10	US-09-900-575-55	Sequence 55, Appl
12	75	46.6	280	10	US-09-900-575-30	Sequence 30, Appl
13	67	41.6	279	10	US-09-900-575-25	Sequence 25, Appl
14	67	41.6	279	10	US-09-900-575-26	Sequence 26, Appl
15	67	41.6	279	10	US-09-900-575-35	Sequence 35, Appl
16	67	41.6	279	10	US-09-900-575-37	Sequence 37, Appl
17	67	41.6	279	10	US-09-900-575-40	Sequence 40, Appl
18	67	41.6	279	10	US-09-900-575-41	Sequence 41, Appl
19	67	41.6	279	10	US-09-900-575-42	Sequence 42, Appl
20	67	41.6	279	10	US-09-900-575-45	Sequence 45, Appl
21	60	37.3	279	10	US-09-900-575-38	Sequence 38, Appl
22	56	34.8	279	10	US-09-900-575-33	Sequence 33, Appl
23	48	29.8	279	10	US-09-900-575-23	Sequence 23, Appl
24	43	26.7	279	10	US-09-900-575-24	Sequence 24, Appl
25	43	26.7	279	10	US-09-900-575-31	Sequence 31, Appl
26	8	5.0	1036	12	US-10-014-882-2	Sequence 2, Appli
27	7	4.3	132	10	US-09-901-436A-11	Sequence 11, Appl
28	7	4.3	1040	10	US-09-912-020-282	Sequence 282, App
29	7	4.3	1040	10	US-09-815-242-10203	Sequence 10203, A
30	6	3.7	40	10	US-09-864-761-37358	Sequence 37358, A
31	6	3.7	46	10	US-09-726-643-65	Sequence 65, Appl
32	6	3.7	54	10	US-09-864-761-34310	Sequence 34310, A
33	6	3.7	62	10	US-09-864-761-35614	Sequence 35614, A
34	6	3.7	93	10	US-09-867-550-1142	Sequence 1142, Ap
35	6	3.7	99	10	US-09-864-761-35148	Sequence 35148, A
36	6	3.7	112	10	US-09-925-300-1428	Sequence 1428, Ap
37	6	3.7	144	10	US-09-728-914-10	Sequence 10, Appl
38	6	3.7	182	10	US-09-864-761-43168	Sequence 43168, A
39	6	3.7	183	10	US-09-739-907-65	Sequence 65, Appl
40	6	3.7	185	12	US-10-036-869-36	Sequence 36, Appl
41	6	3.7	191	10	US-09-998-831-13	Sequence 13, Appl
42	6	3.7	192	10	US-09-925-302-728	Sequence 728, App
43	6	3.7	194	10	US-09-734-017A-74	Sequence 74, Appl
44	6	3.7	213	10	US-09-815-242-13993	Sequence 13993, A
45	6	3.7	252	10	US-09-946-239-10	Sequence 10, Appl
46	6	3.7	275	9	US-09-828-523A-4	Sequence 4, Appli
47	6	3.7	284	9	US-09-828-523A-54	Sequence 54, Appl
48	6	3.7	305	10	US-09-443-704-8	Sequence 8, Appli
49	6	3.7	335	10	US-09-925-300-1575	Sequence 1575, Ap
50	6	3.7	352	10	US-09-771-956-23	Sequence 23, Appl
51	6	3.7	353	10	US-09-962-739-4	Sequence 4, Appli
52	6	3.7	364	10	US-09-922-494-10	Sequence 10, Appl
53	6	3.7	364	10	US-09-922-494-11	Sequence 11, Appl
54	6	3.7	365	10	US-09-925-300-1391	Sequence 1391, Ap
55	6	3.7	372	10	US-09-784-358-6	Sequence 6, Appli
56	6	3.7	376	10	US-09-801-368-206	Sequence 206, App
57	6	3.7	395	10	US-09-771-956-25	Sequence 25, Appl
58	6	3.7	431	10	US-09-835-270-4	Sequence 4, Appli
59	6	3.7	446	10	US-09-784-358-4	Sequence 4, Appli
60	6	3.7	448	10	US-09-939-521-7	Sequence 7, Appli
61	6	3.7	453	12	US-10-083-452-6	Sequence 6, Appli
62	6	3.7	456	10	US-09-962-646-2	Sequence 2, Appli
63	6	3.7	460	9	US-09-813-453A-39	Sequence 39, Appl
64	6	3.7	466	10	US-09-765-205-28	Sequence 28, Appl
65	6	3.7	499	10	US-09-771-956-24	Sequence 24, Appl
66	6	3.7	508	10	US-09-771-956-22	Sequence 22, Appl
67	6	3.7	513	10	US-09-833-745-55	Sequence 55, Appl
68	6	3.7	514	9	US-09-712-363-154	Sequence 154, App
69	6	3.7	518	10	US-09-803-589-10	Sequence 10, Appl
70	6	3.7	523	10	US-09-930-218-4	Sequence 4, Appli
71	6	3.7	541	10	US-09-981-421-4	Sequence 4, Appli
72	6	3.7	541	12	US-10-157-447-2	Sequence 2, Appli
73	6	3.7	551	12	US-10-105-929-16	Sequence 16, Appl
74	6	3.7	576	10	US-09-851-686-6	Sequence 6, Appli
75	6	3.7	592	9	US-09-813-453A-22	Sequence 22, Appl
76	6	3.7	592	9	US-09-813-453A-43	Sequence 43, Appl
77	6	3.7	605	9	US-09-906-209-4	Sequence 4, Appli
78	6	3.7	612	12	US-10-001-851-25	Sequence 25, Appl
79	6	3.7	650	10	US-09-784-358-10	Sequence 10, Appl
80	6	3.7	655	9	US-09-906-209-10	Sequence 10, Appl
81	6	3.7	669	10	US-09-962-739-2	Sequence 2, Appli
82	6	3.7	679	10	US-09-925-302-632	Sequence 632, App
83	6	3.7	724	10	US-09-784-358-8	Sequence 8, Appli
84	6	3.7	727	10	US-09-445-023A-12	Sequence 12, Appl
85	6	3.7	765	10	US-09-975-326-4	Sequence 4, Appli
86	6	3.7	766	10	US-09-975-326-2	Sequence 2, Appli
87	6	3.7	771	10	US-09-784-358-14	Sequence 14, Appl
88	6	3.7	802	10	US-09-838-561-2	Sequence 2, Appli
89	6	3.7	802	10	US-09-816-760-2	Sequence 2, Appli
90	6	3.7	845	10	US-09-784-358-12	Sequence 12, Appl
91	6	3.7	950	10	US-09-321-987B-4	Sequence 4, Appli
92	6	3.7	969	8	US-08-945-749-3	Sequence 3, Appli

93	6	3.7	1374	10	US-09-771-161A-267	Sequence 267, App
94	6	3.7	1404	10	US-09-811-045A-1	Sequence 1, Appli
95	6	3.7	1617	10	US-09-784-358-16	Sequence 16, Appl
96	6	3.7	1653	10	US-09-741-669-402	Sequence 402, App
97	6	3.7	1691	10	US-09-784-358-2	Sequence 2, Appli
98	6	3.7	2121	10	US-09-995-542-3	Sequence 3, Appli
99	6	3.7	2167	10	US-09-995-542-2	Sequence 2, Appli
100	6	3.7	2209	9	US-09-902-941-1903	Sequence 1903, Ap
101	6	3.7	3056	8	US-08-984-090-2	Sequence 2, Appli
102	6	3.7	5215	9	US-09-860-846-2	Sequence 2, Appli
103	6	3.7	5215	10	US-09-861-289-2	Sequence 2, Appli
104	5	3.1	5	10	US-09-751-100B-35	Sequence 35, Appl
105	5	3.1	9	10	US-09-808-037-17	Sequence 17, Appl
106	5	3.1	9	10	US-09-808-037-19	Sequence 19, Appl
107	5	3.1	9	10	US-09-976-787-6	Sequence 6, Appli
108	5	3.1	9	10	US-09-865-198-6	Sequence 6, Appli
109	5	3.1	13	10	US-09-766-412-44	Sequence 44, Appl
110	5	3.1	15	9	US-10-072-841-43	Sequence 43, Appl
111	5	3.1	16	9	US-10-029-009-13	Sequence 13, Appl
112	5	3.1	17	9	US-10-029-009-25	Sequence 25, Appl
113	5	3.1	19	10	US-09-864-761-48306	Sequence 48306, A
114	5	3.1	22	9	US-09-991-548-22	Sequence 22, Appl
115	5	3.1	22	9	US-09-899-235-20	Sequence 20, Appl
116	5	3.1	23	10	US-09-864-761-36822	Sequence 36822, A
117	5	3.1	24	10	US-09-030-619-213	Sequence 213, App
118	5	3.1	24	10	US-09-864-761-44483	Sequence 44483, A
119	5	3.1	24	10	US-09-917-340-2	Sequence 2, Appli
120	5	3.1	26	10	US-09-099-823-25	Sequence 25, Appl
121	5	3.1	26	10	US-09-864-761-43769	Sequence 43769, A
122	5	3.1	26	10	US-09-864-761-47278	Sequence 47278, A
123	5	3.1	27	10	US-09-864-761-37931	Sequence 37931, A
124	5	3.1	27	10	US-09-864-761-42537	Sequence 42537, A
125	5	3.1	27	10	US-09-938-700-12	Sequence 12, Appl
126	5	3.1	28	8	US-08-424-550B-573	Sequence 573, App
127	5	3.1	28	10	US-09-938-700-5	Sequence 5, Appli
128	5	3.1	29	10	US-09-864-761-41724	Sequence 41724, A
129	5	3.1	29	10	US-09-935-291A-35	Sequence 35, Appl
130	5	3.1	29	10	US-09-974-449-31	Sequence 31, Appl
131	5	3.1	30	10	US-09-864-761-33704	Sequence 33704, A
132	5	3.1	30	10	US-09-864-761-36699	Sequence 36699, A
133	5	3.1	30	10	US-09-925-299-1432	Sequence 1432, Ap
134	5	3.1	30	10	US-09-908-322-50	Sequence 50, Appl
135	5	3.1	31	10	US-09-938-700-2	Sequence 2, Appli
136	5	3.1	32	10	US-09-864-761-34552	Sequence 34552, A
137	5	3.1	33	10	US-09-864-761-45937	Sequence 45937, A
138	5	3.1	33	10	US-09-864-761-47894	Sequence 47894, A
139	5	3.1	33	10	US-09-938-700-9	Sequence 9, Appli
140	5	3.1	34	10	US-09-281-717-20	Sequence 20, Appl
141	5	3.1	35	10	US-09-864-761-33420	Sequence 33420, A
142	5	3.1	35	10	US-09-864-761-44044	Sequence 44044, A
143	5	3.1	36	10	US-09-864-761-37974	Sequence 37974, A
144	5	3.1	37	10	US-09-205-658-100	Sequence 100, App
145	5	3.1	38	10	US-09-864-761-36437	Sequence 36437, A
146	5	3.1	38	10	US-09-864-761-36955	Sequence 36955, A
147	5	3.1	41	10	US-09-864-761-34825	Sequence 34825, A
148	5	3.1	41	10	US-09-864-761-41709	Sequence 41709, A
149	5	3.1	42	10	US-09-864-761-35422	Sequence 35422, A
150	5	3.1	42	10	US-09-864-761-36527	Sequence 36527, A
151	5	3.1	42	10	US-09-864-761-47093	Sequence 47093, A
152	5	3.1	43	12	US-10-042-417-79	Sequence 79, Appl
153	5	3.1	44	10	US-09-881-752A-292	Sequence 292, App
154	5	3.1	46	10	US-09-205-658-83	Sequence 83, Appl
155	5	3.1	46	10	US-09-844-353A-83	Sequence 83, Appl
156	5	3.1	46	10	US-09-864-761-36363	Sequence 36363, A
157	5	3.1	47	10	US-09-864-761-34991	Sequence 34991, A
158	5	3.1	47	10	US-09-864-761-43524	Sequence 43524, A
159	5	3.1	48	12	US-10-001-879-147	Sequence 147, App
160	5	3.1	49	10	US-09-864-761-40743	Sequence 40743, A
161	5	3.1	49	10	US-09-925-301-1678	Sequence 1678, Ap
162	5	3.1	50	10	US-09-864-761-46350	Sequence 46350, A
163	5	3.1	50	12	US-10-007-331-1	Sequence 1, Appli
164	5	3.1	51	10	US-09-864-761-40434	Sequence 40434, A
165	5	3.1	51	10	US-09-864-761-40652	Sequence 40652, A
166	5	3.1	51	10	US-09-764-869-870	Sequence 870, App
167	5	3.1	53	10	US-09-864-761-36488	Sequence 36488, A
168	5	3.1	53	10	US-09-864-761-37908	Sequence 37908, A
169	5	3.1	53	10	US-09-867-550-1076	Sequence 1076, Ap
170	5	3.1	57	10	US-09-864-761-42965	Sequence 42965, A
171	5	3.1	58	10	US-09-864-761-35371	Sequence 35371, A
172	5	3.1	58	10	US-09-864-761-36592	Sequence 36592, A
173	5	3.1	59	10	US-09-840-787-3	Sequence 3, Appli
174	5	3.1	59	12	US-10-001-879-166	Sequence 166, App
175	5	3.1	60	10	US-09-864-761-44295	Sequence 44295, A
176	5	3.1	61	10	US-09-864-761-38561	Sequence 38561, A
177	5	3.1	62	10	US-09-739-907-98	Sequence 98, Appl
178	5	3.1	63	10	US-09-764-869-630	Sequence 630, App
179	5	3.1	63	10	US-09-948-080-18	Sequence 18, Appl
180	5	3.1	65	10	US-09-864-761-34462	Sequence 34462, A
181	5	3.1	66	10	US-09-764-887-265	Sequence 265, App
182	5	3.1	66	10	US-09-925-300-1832	Sequence 1832, Ap
183	5	3.1	67	9	US-10-002-344A-151	Sequence 151, App
184	5	3.1	68	10	US-09-764-877-1714	Sequence 1714, Ap
185	5	3.1	69	10	US-09-864-761-36253	Sequence 36253, A
186	5	3.1	70	9	US-09-981-876-269	Sequence 269, App
187	5	3.1	70	9	US-09-792-793A-91	Sequence 91, Appl
188	5	3.1	70	10	US-09-867-550-934	Sequence 934, App
189	5	3.1	71	10	US-09-764-847-926	Sequence 926, App
190	5	3.1	72	10	US-09-864-761-47147	Sequence 47147, A
191	5	3.1	72	10	US-09-731-872-382	Sequence 382, App
192	5	3.1	75	10	US-09-864-761-46024	Sequence 46024, A
193	5	3.1	77	9	US-09-852-797-67	Sequence 67, Appl
194	5	3.1	77	10	US-09-925-301-1610	Sequence 1610, Ap
195	5	3.1	77	10	US-09-853-161-67	Sequence 67, Appl
196	5	3.1	77	10	US-09-852-659A-67	Sequence 67, Appl
197	5	3.1	78	10	US-09-864-761-44201	Sequence 44201, A
198	5	3.1	81	9	US-10-103-511-10	Sequence 10, Appl
199	5	3.1	81	10	US-09-864-761-43368	Sequence 43368, A
200	5	3.1	81	10	US-09-805-204-10	Sequence 10, Appl
201	5	3.1	83	10	US-09-864-761-44350	Sequence 44350, A
202	5	3.1	83	10	US-09-864-761-48951	Sequence 48951, A
203	5	3.1	84	10	US-09-815-242-13034	Sequence 13034, A
204	5	3.1	86	10	US-09-764-869-669	Sequence 669, App
205	5	3.1	86	10	US-09-764-860-544	Sequence 544, App
206	5	3.1	86	10	US-09-731-872-290	Sequence 290, App
207	5	3.1	87	10	US-09-925-300-990	Sequence 990, App
208	5	3.1	89	10	US-09-860-232A-9	Sequence 9, Appli
209	5	3.1	89	10	US-09-864-761-42696	Sequence 42696, A
210	5	3.1	90	10	US-09-864-761-34584	Sequence 34584, A
211	5	3.1	90	10	US-09-864-761-46461	Sequence 46461, A
212	5	3.1	91	10	US-09-864-761-36976	Sequence 36976, A
213	5	3.1	91	10	US-09-864-761-43033	Sequence 43033, A
214	5	3.1	92	9	US-09-981-876-238	Sequence 238, App
215	5	3.1	93	9	US-09-981-876-183	Sequence 183, App
216	5	3.1	93	10	US-09-764-877-1262	Sequence 1262, Ap
217	5	3.1	94	10	US-09-864-761-48196	Sequence 48196, A
218	5	3.1	95	9	US-10-002-344A-167	Sequence 167, App
219	5	3.1	96	10	US-09-905-243-17	Sequence 17, Appl
220	5	3.1	96	10	US-09-764-847-792	Sequence 792, App
221	5	3.1	97	9	US-09-895-913A-58	Sequence 58, Appl
222	5	3.1	97	9	US-09-764-868-1171	Sequence 1171, Ap
223	5	3.1	99	10	US-09-925-299-1482	Sequence 1482, Ap
224	5	3.1	100	9	US-09-712-363-272	Sequence 272, App
225	5	3.1	100	10	US-09-815-242-11047	Sequence 11047, A
226	5	3.1	101	9	US-09-464-099A-17	Sequence 17, Appl
227	5	3.1	101	10	US-09-861-696-17	Sequence 17, Appl
228	5	3.1	101	10	US-09-940-063-8	Sequence 8, Appli
229	5	3.1	101	10	US-09-854-731-2	Sequence 2, Appli
230	5	3.1	102	10	US-09-864-761-43840	Sequence 43840, A
231	5	3.1	103	10	US-09-815-242-11108	Sequence 11108, A
232	5	3.1	104	10	US-09-764-846-153	Sequence 153, App
233	5	3.1	105	10	US-09-764-877-1972	Sequence 1972, Ap
234	5	3.1	106	10	US-09-864-761-33300	Sequence 33300, A
235	5	3.1	106	10	US-09-976-787-24	Sequence 24, Appl
236	5	3.1	106	10	US-09-865-198-23	Sequence 23, Appl
237	5	3.1	107	9	US-09-144-886-75	Sequence 75, Appl
238	5	3.1	107	9	US-09-986-234-4	Sequence 4, Appli

239	5	3.1	108	10	US-09-742-858-2	Sequence 2, Appli
240	5	3.1	108	10	US-09-742-858-4	Sequence 4, Appli
241	5	3.1	108	10	US-09-822-263-14	Sequence 14, Appl
242	5	3.1	108	10	US-09-976-787-8	Sequence 8, Appli
243	5	3.1	108	10	US-09-865-198-8	Sequence 8, Appli
244	5	3.1	108	10	US-09-957-943-14	Sequence 14, Appl
245	5	3.1	108	10	US-09-747-835A-6	Sequence 6, Appli
246	5	3.1	109	10	US-09-802-077-1	Sequence 1, Appli
247	5	3.1	109	10	US-09-802-096-1	Sequence 1, Appli
248	5	3.1	109	10	US-09-934-586A-6	Sequence 6, Appli
249	5	3.1	109	10	US-09-814-122-43	Sequence 43, Appl
250	5	3.1	109	10	US-09-867-550-106	Sequence 106, App
251	5	3.1	110	10	US-09-867-550-822	Sequence 822, App
252	5	3.1	112	8	US-08-424-550B-68	Sequence 68, Appl
253	5	3.1	112	10	US-09-864-761-35620	Sequence 35620, A
254	5	3.1	113	10	US-09-867-550-70	Sequence 70, Appl
255	5	3.1	113	10	US-09-867-550-1398	Sequence 1398, Ap
256	5	3.1	116	10	US-09-942-446-5	Sequence 5, Appli
257	5	3.1	116	10	US-09-939-980-304	Sequence 304, App
258	5	3.1	116	10	US-09-938-970-5	Sequence 5, Appli
259	5	3.1	116	10	US-09-731-872-458	Sequence 458, App
260	5	3.1	117	9	US-09-293-854-4	Sequence 4, Appli
261	5	3.1	117	10	US-09-815-242-11223	Sequence 11223, A
262	5	3.1	118	10	US-09-912-020-394	Sequence 394, App
263	5	3.1	118	10	US-09-925-301-1091	Sequence 1091, Ap
264	5	3.1	118	10	US-09-764-869-626	Sequence 626, App
265	5	3.1	118	10	US-09-815-242-10176	Sequence 10176, A
266	5	3.1	118	10	US-09-815-242-11882	Sequence 11882, A
267	5	3.1	118	10	US-09-815-242-13941	Sequence 13941, A
268	5	3.1	119	10	US-09-864-761-36209	Sequence 36209, A
269	5	3.1	119	10	US-09-808-037-28	Sequence 28, Appl
270	5	3.1	120	10	US-09-864-761-47331	Sequence 47331, A
271	5	3.1	120	10	US-09-764-877-1108	Sequence 1108, Ap
272	5	3.1	120	10	US-09-840-459-85	Sequence 85, Appl
273	5	3.1	121	10	US-09-864-761-40769	Sequence 40769, A
274	5	3.1	122	10	US-09-893-737-34	Sequence 34, Appl
275	5	3.1	122	12	US-10-042-417-38	Sequence 38, Appl
276	5	3.1	124	10	US-09-764-878-163	Sequence 163, App
277	5	3.1	125	10	US-09-764-887-262	Sequence 262, App
278	5	3.1	126	9	US-09-960-631A-7	Sequence 7, Appli
279	5	3.1	126	10	US-09-864-761-34753	Sequence 34753, A
280	5	3.1	127	10	US-09-864-761-38565	Sequence 38565, A
281	5	3.1	128	10	US-09-864-761-37894	Sequence 37894, A
282	5	3.1	128	10	US-09-925-300-1730	Sequence 1730, Ap
283	5	3.1	129	10	US-09-730-617-90	Sequence 90, Appl
284	5	3.1	130	10	US-09-780-224A-9	Sequence 9, Appli
285	5	3.1	130	10	US-09-731-872-289	Sequence 289, App
286	5	3.1	131	10	US-09-881-823-6	Sequence 6, Appli
287	5	3.1	133	10	US-09-798-029-25	Sequence 25, Appl
288	5	3.1	134	10	US-09-925-302-694	Sequence 694, App
289	5	3.1	135	10	US-09-864-761-47958	Sequence 47958, A
290	5	3.1	137	9	US-09-764-868-734	Sequence 734, App
291	5	3.1	137	10	US-09-733-507-14	Sequence 14, Appl
292	5	3.1	137	10	US-09-908-711-94	Sequence 94, Appl
293	5	3.1	137	10	US-09-864-761-45751	Sequence 45751, A
294	5	3.1	139	9	US-10-032-159A-16	Sequence 16, Appl
295	5	3.1	140	12	US-10-116-378-30	Sequence 30, Appl
296	5	3.1	141	10	US-09-877-156-22	Sequence 22, Appl
297	5	3.1	141	10	US-09-771-161A-99	Sequence 99, Appl
298	5	3.1	142	10	US-09-864-761-42510	Sequence 42510, A
299	5	3.1	142	10	US-09-738-973-213	Sequence 213, App
300	5	3.1	143	9	US-09-779-050A-8	Sequence 8, Appli
301	5	3.1	143	9	US-09-779-050A-9	Sequence 9, Appli
302	5	3.1	143	9	US-09-852-797-107	Sequence 107, App
303	5	3.1	143	10	US-09-864-761-34717	Sequence 34717, A
304	5	3.1	143	10	US-09-925-299-1103	Sequence 1103, Ap
305	5	3.1	143	10	US-09-853-161-107	Sequence 107, App
306	5	3.1	143	10	US-09-852-659A-107	Sequence 107, App
307	5	3.1	145	9	US-09-860-670-126	Sequence 126, App
308	5	3.1	145	9	US-09-764-868-1176	Sequence 1176, Ap
309	5	3.1	145	10	US-09-764-853-626	Sequence 626, App
310	5	3.1	146	10	US-09-925-299-833	Sequence 833, App
311	5	3.1	146	10	US-09-925-299-1075	Sequence 1075, Ap
312	5	3.1	146	12	US-10-001-843-193	Sequence 193, App
313	5	3.1	147	9	US-09-756-854-19	Sequence 19, Appl
314	5	3.1	147	9	US-10-041-574-19	Sequence 19, Appl
315	5	3.1	148	10	US-09-925-300-1770	Sequence 1770, Ap
316	5	3.1	149	10	US-09-925-299-895	Sequence 895, App
317	5	3.1	149	10	US-09-934-465-17	Sequence 17, Appl
318	5	3.1	153	9	US-10-077-040-3	Sequence 3, Appli
319	5	3.1	153	9	US-09-992-598-334	Sequence 334, App
320	5	3.1	153	10	US-09-989-722-334	Sequence 334, App
321	5	3.1	153	10	US-09-989-723-334	Sequence 334, App
322	5	3.1	153	10	US-09-989-279-334	Sequence 334, App
323	5	3.1	153	10	US-09-989-727-334	Sequence 334, App
324	5	3.1	153	10	US-09-989-731-334	Sequence 334, App
325	5	3.1	153	10	US-09-989-732-334	Sequence 334, App
326	5	3.1	153	10	US-09-991-073-334	Sequence 334, App
327	5	3.1	153	10	US-09-990-442-334	Sequence 334, App
328	5	3.1	153	10	US-09-991-163-334	Sequence 334, App
329	5	3.1	153	10	US-09-993-604-334	Sequence 334, App
330	5	3.1	153	10	US-09-990-456-334	Sequence 334, App
331	5	3.1	153	10	US-09-989-721-334	Sequence 334, App
332	5	3.1	153	12	US-10-052-586-232	Sequence 232, App
333	5	3.1	154	10	US-09-764-864-1004	Sequence 1004, Ap
334	5	3.1	155	10	US-09-929-493-23	Sequence 23, Appl
335	5	3.1	155	10	US-09-952-432A-2	Sequence 2, Appli
336	5	3.1	155	10	US-09-952-432A-16	Sequence 16, Appl
337	5	3.1	156	10	US-09-864-761-48065	Sequence 48065, A
338	5	3.1	156	10	US-09-815-242-10329	Sequence 10329, A
339	5	3.1	157	9	US-09-992-598-103	Sequence 103, App
340	5	3.1	157	9	US-09-764-868-913	Sequence 913, App
341	5	3.1	157	10	US-09-989-722-103	Sequence 103, App
342	5	3.1	157	10	US-09-989-723-103	Sequence 103, App
343	5	3.1	157	10	US-09-989-279-103	Sequence 103, App
344	5	3.1	157	10	US-09-989-727-103	Sequence 103, App
345	5	3.1	157	10	US-09-867-550-656	Sequence 656, App
346	5	3.1	157	10	US-09-989-731-103	Sequence 103, App
347	5	3.1	157	10	US-09-989-732-103	Sequence 103, App
348	5	3.1	157	10	US-09-991-073-103	Sequence 103, App
349	5	3.1	157	10	US-09-990-442-103	Sequence 103, App
350	5	3.1	157	10	US-09-991-163-103	Sequence 103, App
351	5	3.1	157	10	US-09-993-604-103	Sequence 103, App
352	5	3.1	157	10	US-09-862-658-6	Sequence 6, Appli
353	5	3.1	157	10	US-09-990-456-103	Sequence 103, App
354	5	3.1	157	10	US-09-989-721-103	Sequence 103, App
355	5	3.1	157	12	US-10-052-586-90	Sequence 90, Appl
356	5	3.1	158	9	US-09-887-593-2	Sequence 2, Appli
357	5	3.1	158	10	US-09-731-872-260	Sequence 260, App
358	5	3.1	158	12	US-10-052-586-6	Sequence 6, Appli
359	5	3.1	159	10	US-09-764-864-1193	Sequence 1193, Ap
360	5	3.1	161	10	US-09-864-761-40654	Sequence 40654, A
361	5	3.1	162	10	US-09-764-864-1180	Sequence 1180, Ap
362	5	3.1	163	10	US-09-939-980-420	Sequence 420, App
363	5	3.1	163	10	US-09-764-864-1440	Sequence 1440, Ap
364	5	3.1	164	9	US-09-735-713A-6	Sequence 6, Appli
365	5	3.1	166	10	US-09-925-299-1086	Sequence 1086, Ap
366	5	3.1	166	10	US-09-747-835A-56	Sequence 56, Appl
367	5	3.1	168	9	US-09-770-528-13	Sequence 13, Appl
368	5	3.1	168	10	US-09-815-242-10995	Sequence 10995, A
369	5	3.1	168	10	US-09-159-469-37	Sequence 37, Appl
370	5	3.1	168	10	US-09-798-042-37	Sequence 37, Appl
371	5	3.1	169	9	US-09-770-528-15	Sequence 15, Appl
372	5	3.1	169	10	US-09-925-302-633	Sequence 633, App
373	5	3.1	169	10	US-09-925-301-1172	Sequence 1172, Ap
374	5	3.1	169	10	US-09-925-299-1125	Sequence 1125, Ap
375	5	3.1	169	10	US-09-775-046-4	Sequence 4, Appli
376	5	3.1	170	10	US-09-099-823-22	Sequence 22, Appl
377	5	3.1	170	10	US-09-951-845-2	Sequence 2, Appli
378	5	3.1	171	10	US-09-775-925-30	Sequence 30, Appl
379	5	3.1	171	10	US-09-864-761-34202	Sequence 34202, A
380	5	3.1	171	10	US-09-847-519A-15	Sequence 15, Appl
381	5	3.1	172	10	US-09-800-729-138	Sequence 138, App
382	5	3.1	172	10	US-09-764-853-627	Sequence 627, App
383	5	3.1	172	10	US-09-764-898-251	Sequence 251, App
384	5	3.1	172	10	US-09-764-903-49	Sequence 49, Appl

385	5	3.1	173	10	US-09-730-617-59	Sequence 59, Appl
386	5	3.1	173	10	US-09-764-898-178	Sequence 178, App
387	5	3.1	174	10	US-09-764-853-563	Sequence 563, App
388	5	3.1	175	10	US-09-814-122-68	Sequence 68, Appl
389	5	3.1	176	10	US-09-976-451-8	Sequence 8, Appli
390	5	3.1	176	10	US-09-925-300-1547	Sequence 1547, Ap
391	5	3.1	177	10	US-09-746-801A-53	Sequence 53, Appl
392	5	3.1	178	10	US-09-828-644-95	Sequence 95, Appl
393	5	3.1	178	10	US-09-925-301-1390	Sequence 1390, Ap
394	5	3.1	178	10	US-09-815-242-5866	Sequence 5866, Ap
395	5	3.1	178	10	US-09-815-242-10448	Sequence 10448, A
396	5	3.1	178	10	US-09-815-242-12220	Sequence 12220, A
397	5	3.1	178	10	US-09-815-242-12830	Sequence 12830, A
398	5	3.1	178	10	US-09-970-711-2	Sequence 2, Appli
399	5	3.1	180	10	US-09-739-254-78	Sequence 78, Appl
400	5	3.1	180	10	US-09-904-615-78	Sequence 78, Appl
401	5	3.1	181	10	US-09-925-302-515	Sequence 515, App
402	5	3.1	183	10	US-09-764-870-516	Sequence 516, App
403	5	3.1	183	10	US-09-731-872-420	Sequence 420, App
404	5	3.1	186	9	US-09-989-350-6	Sequence 6, Appli
405	5	3.1	187	9	US-09-992-598-189	Sequence 189, App
406	5	3.1	187	10	US-09-811-118-1	Sequence 1, Appli
407	5	3.1	187	10	US-09-989-722-189	Sequence 189, App
408	5	3.1	187	10	US-09-989-723-189	Sequence 189, App
409	5	3.1	187	10	US-09-989-279-189	Sequence 189, App
410	5	3.1	187	10	US-09-989-727-189	Sequence 189, App
411	5	3.1	187	10	US-09-989-731-189	Sequence 189, App
412	5	3.1	187	10	US-09-989-732-189	Sequence 189, App
413	5	3.1	187	10	US-09-991-073-189	Sequence 189, App
414	5	3.1	187	10	US-09-990-442-189	Sequence 189, App
415	5	3.1	187	10	US-09-991-163-189	Sequence 189, App
416	5	3.1	187	10	US-09-993-604-189	Sequence 189, App
417	5	3.1	187	10	US-09-990-456-189	Sequence 189, App
418	5	3.1	187	10	US-09-989-721-189	Sequence 189, App
419	5	3.1	187	12	US-10-052-586-426	Sequence 426, App
420	5	3.1	192	10	US-09-823-069-4	Sequence 4, Appli
421	5	3.1	192	10	US-09-766-378A-38	Sequence 38, Appl
422	5	3.1	193	10	US-09-729-674-170	Sequence 170, App
423	5	3.1	196	10	US-09-925-299-1008	Sequence 1008, Ap
424	5	3.1	196	10	US-09-764-846-201	Sequence 201, App
425	5	3.1	202	10	US-09-960-253-184	Sequence 184, App
426	5	3.1	203	10	US-09-815-242-11474	Sequence 11474, A
427	5	3.1	203	10	US-09-809-545A-2	Sequence 2, Appli
428	5	3.1	203	10	US-09-788-654A-2	Sequence 2, Appli
429	5	3.1	203	10	US-09-972-529-2	Sequence 2, Appli
430	5	3.1	205	10	US-09-734-017A-22	Sequence 22, Appl
431	5	3.1	206	12	US-10-062-254-128	Sequence 128, App
432	5	3.1	207	9	US-09-984-245-201	Sequence 201, App
433	5	3.1	207	10	US-09-822-485-19	Sequence 19, Appl
434	5	3.1	207	10	US-09-750-963-6	Sequence 6, Appli
435	5	3.1	207	10	US-09-805-805-5	Sequence 5, Appli
436	5	3.1	207	10	US-09-805-805-7	Sequence 7, Appli
437	5	3.1	207	10	US-09-817-814-24	Sequence 24, Appl
438	5	3.1	207	12	US-10-005-646-6	Sequence 6, Appli
439	5	3.1	208	10	US-09-815-242-5302	Sequence 5302, Ap
440	5	3.1	210	9	US-09-855-835-2	Sequence 2, Appli
441	5	3.1	210	10	US-09-925-300-1354	Sequence 1354, Ap
442	5	3.1	211	10	US-09-912-020-296	Sequence 296, App
443	5	3.1	212	10	US-09-730-617-60	Sequence 60, Appl
444	5	3.1	212	10	US-09-730-617-78	Sequence 78, Appl
445	5	3.1	213	10	US-09-815-242-10129	Sequence 10129, A
446	5	3.1	213	10	US-09-815-242-14070	Sequence 14070, A
447	5	3.1	215	10	US-09-894-998-46	Sequence 46, Appl
448	5	3.1	216	10	US-09-864-761-33387	Sequence 33387, A
449	5	3.1	216	10	US-09-908-180-4	Sequence 4, Appli
450	5	3.1	216	10	US-09-747-155-7	Sequence 7, Appli
451	5	3.1	216	10	US-09-747-155-373	Sequence 373, App
452	5	3.1	217	10	US-09-925-301-996	Sequence 996, App
453	5	3.1	217	10	US-09-925-301-1073	Sequence 1073, Ap
454	5	3.1	219	10	US-09-815-242-12168	Sequence 12168, A
455	5	3.1	219	10	US-09-815-242-13043	Sequence 13043, A
456	5	3.1	221	9	US-09-971-536-48	Sequence 48, Appl
457	5	3.1	221	10	US-09-925-302-716	Sequence 716, App
458	5	3.1	221	10	US-09-764-864-1480	Sequence 1480, Ap
459	5	3.1	222	9	US-09-764-868-1112	Sequence 1112, Ap
460	5	3.1	222	10	US-09-809-746-2	Sequence 2, Appli
461	5	3.1	222	10	US-09-864-761-42926	Sequence 42926, A
462	5	3.1	222	10	US-09-893-737-108	Sequence 108, App
463	5	3.1	223	9	US-09-989-545-20	Sequence 20, Appl
464	5	3.1	223	10	US-09-764-870-436	Sequence 436, App
465	5	3.1	225	9	US-09-764-868-692	Sequence 692, App
466	5	3.1	225	10	US-09-815-242-12109	Sequence 12109, A
467	5	3.1	225	10	US-09-815-242-13706	Sequence 13706, A
468	5	3.1	229	10	US-09-350-874-70	Sequence 70, Appl
469	5	3.1	229	10	US-09-765-213A-8	Sequence 8, Appli
470	5	3.1	230	10	US-09-864-761-34040	Sequence 34040, A
471	5	3.1	230	10	US-09-815-242-5661	Sequence 5661, Ap
472	5	3.1	231	10	US-09-765-272-176	Sequence 176, App
473	5	3.1	231	10	US-09-840-795-19	Sequence 19, Appl
474	5	3.1	233	9	US-09-972-970-4	Sequence 4, Appli
475	5	3.1	233	12	US-10-062-254-202	Sequence 202, App
476	5	3.1	236	10	US-09-910-150-33	Sequence 33, Appl
477	5	3.1	237	10	US-09-746-284-3	Sequence 3, Appli
478	5	3.1	238	10	US-09-976-787-29	Sequence 29, Appl
479	5	3.1	238	10	US-09-865-198-28	Sequence 28, Appl
480	5	3.1	239	10	US-09-741-669-310	Sequence 310, App
481	5	3.1	239	10	US-09-912-020-287	Sequence 287, App
482	5	3.1	239	10	US-09-808-037-6	Sequence 6, Appli
483	5	3.1	240	10	US-09-976-787-28	Sequence 28, Appl
484	5	3.1	240	10	US-09-865-198-27	Sequence 27, Appl
485	5	3.1	241	10	US-09-925-297-542	Sequence 542, App
486	5	3.1	242	10	US-09-764-864-1341	Sequence 1341, Ap
487	5	3.1	243	9	US-09-984-245-343	Sequence 343, App
488	5	3.1	244	9	US-10-002-344A-276	Sequence 276, App
489	5	3.1	244	10	US-09-939-980-379	Sequence 379, App
490	5	3.1	244	10	US-09-764-864-1075	Sequence 1075, Ap
491	5	3.1	245	10	US-09-925-297-619	Sequence 619, App
492	5	3.1	245	10	US-09-925-297-624	Sequence 624, App
493	5	3.1	246	10	US-09-864-761-36064	Sequence 36064, A
494	5	3.1	246	12	US-10-062-254-218	Sequence 218, App
495	5	3.1	247	10	US-09-923-779-154	Sequence 154, App
496	5	3.1	247	10	US-09-800-103-32	Sequence 32, Appl
497	5	3.1	247	10	US-09-764-864-805	Sequence 805, App
498	5	3.1	248	9	US-09-925-664-27	Sequence 27, Appl
499	5	3.1	249	8	US-08-818-581B-10	Sequence 10, Appl
500	5	3.1	252	10	US-09-867-550-906	Sequence 906, App
501	5	3.1	252	10	US-09-927-616A-9	Sequence 9, Appli
502	5	3.1	255	9	US-10-027-199-2	Sequence 2, Appli
503	5	3.1	255	10	US-09-739-394-9	Sequence 9, Appli
504	5	3.1	255	10	US-09-826-212-11	Sequence 11, Appl
505	5	3.1	255	10	US-09-925-299-894	Sequence 894, App
506	5	3.1	255	10	US-09-935-727-13	Sequence 13, Appl
507	5	3.1	255	12	US-10-097-330-9	Sequence 9, Appli
508	5	3.1	256	10	US-09-950-368-12	Sequence 12, Appl
509	5	3.1	257	10	US-09-764-898-288	Sequence 288, App
510	5	3.1	258	9	US-09-712-363-220	Sequence 220, App
511	5	3.1	258	10	US-09-994-485-16	Sequence 16, Appl
512	5	3.1	260	10	US-09-925-299-919	Sequence 919, App
513	5	3.1	260	10	US-09-815-242-11638	Sequence 11638, A
514	5	3.1	261	9	US-10-047-260-6	Sequence 6, Appli
515	5	3.1	261	10	US-09-864-761-43223	Sequence 43223, A
516	5	3.1	261	10	US-09-940-037A-29	Sequence 29, Appl
517	5	3.1	262	9	US-09-813-453A-11	Sequence 11, Appl
518	5	3.1	263	10	US-09-811-672-10	Sequence 10, Appl
519	5	3.1	266	10	US-09-876-225-3	Sequence 3, Appli
520	5	3.1	266	10	US-09-911-826A-6	Sequence 6, Appli
521	5	3.1	267	10	US-09-864-761-33871	Sequence 33871, A
522	5	3.1	267	10	US-09-864-761-34858	Sequence 34858, A
523	5	3.1	267	10	US-09-921-232-17	Sequence 17, Appl
524	5	3.1	267	10	US-09-921-330-17	Sequence 17, Appl
525	5	3.1	267	10	US-09-921-329-17	Sequence 17, Appl
526	5	3.1	268	9	US-09-998-582-2	Sequence 2, Appli
527	5	3.1	268	10	US-09-925-637-60	Sequence 60, Appl
528	5	3.1	269	10	US-09-864-761-38036	Sequence 38036, A
529	5	3.1	269	10	US-09-925-297-576	Sequence 576, App
530	5	3.1	270	9	US-09-934-268-2	Sequence 2, Appli

531	5	3.1	270	10	US-09-191-687B-6	Sequence 6, Appli
532	5	3.1	270	10	US-09-923-779-152	Sequence 152, App
533	5	3.1	271	10	US-09-925-299-978	Sequence 978, App
534	5	3.1	272	9	US-09-989-350-8	Sequence 8, Appli
535	5	3.1	272	10	US-09-815-242-5738	Sequence 5738, Ap
536	5	3.1	272	10	US-09-925-300-1450	Sequence 1450, Ap
537	5	3.1	272	12	US-10-001-913-1	Sequence 1, Appli
538	5	3.1	273	10	US-09-925-297-695	Sequence 695, App
539	5	3.1	275	10	US-09-815-242-12069	Sequence 12069, A
540	5	3.1	275	10	US-09-754-997A-6	Sequence 6, Appli
541	5	3.1	276	10	US-09-881-752A-100	Sequence 100, App
542	5	3.1	277	10	US-09-839-339A-1	Sequence 1, Appli
543	5	3.1	278	10	US-09-246-129B-6	Sequence 6, Appli
544	5	3.1	278	10	US-09-899-059-6	Sequence 6, Appli
545	5	3.1	279	8	US-08-971-317A-4	Sequence 4, Appli
546	5	3.1	279	9	US-09-960-631A-2	Sequence 2, Appli
547	5	3.1	279	9	US-10-017-910-5	Sequence 5, Appli
548	5	3.1	279	10	US-09-193-663-4	Sequence 4, Appli
549	5	3.1	279	10	US-09-861-451A-28	Sequence 28, Appl
550	5	3.1	279	10	US-09-832-197-4	Sequence 4, Appli
551	5	3.1	280	9	US-09-960-631A-5	Sequence 5, Appli
552	5	3.1	280	9	US-09-712-363-216	Sequence 216, App
553	5	3.1	280	10	US-09-815-242-10821	Sequence 10821, A
554	5	3.1	281	8	US-08-971-317A-6	Sequence 6, Appli
555	5	3.1	281	8	US-08-818-581B-8	Sequence 8, Appli
556	5	3.1	281	10	US-09-802-669-25	Sequence 25, Appl
557	5	3.1	281	10	US-09-925-302-498	Sequence 498, App
558	5	3.1	281	10	US-09-193-663-6	Sequence 6, Appli
559	5	3.1	281	10	US-09-815-242-12170	Sequence 12170, A
560	5	3.1	281	10	US-09-815-242-13016	Sequence 13016, A
561	5	3.1	281	10	US-09-027-287-6	Sequence 6, Appli
562	5	3.1	281	10	US-09-252-656B-6	Sequence 6, Appli
563	5	3.1	281	10	US-09-929-493-6	Sequence 6, Appli
564	5	3.1	281	10	US-09-927-110-1	Sequence 1, Appli
565	5	3.1	281	12	US-10-012-452-13	Sequence 13, Appl
566	5	3.1	283	10	US-09-815-242-13476	Sequence 13476, A
567	5	3.1	284	10	US-09-800-103-8	Sequence 8, Appli
568	5	3.1	284	10	US-09-800-103-20	Sequence 20, Appl
569	5	3.1	284	12	US-10-041-400A-7	Sequence 7, Appli
570	5	3.1	284	12	US-10-041-264A-7	Sequence 7, Appli
571	5	3.1	284	12	US-10-042-091A-7	Sequence 7, Appli
572	5	3.1	285	9	US-10-117-824-3	Sequence 3, Appli
573	5	3.1	287	10	US-09-415-540-5	Sequence 5, Appli
574	5	3.1	288	10	US-09-778-971-6	Sequence 6, Appli
575	5	3.1	289	9	US-09-970-989-10	Sequence 10, Appl
576	5	3.1	289	10	US-09-815-242-13678	Sequence 13678, A
577	5	3.1	290	9	US-10-041-006A-7	Sequence 7, Appli
578	5	3.1	290	10	US-09-800-103-10	Sequence 10, Appl
579	5	3.1	290	10	US-09-800-103-22	Sequence 22, Appl
580	5	3.1	290	10	US-09-903-814-16	Sequence 16, Appl
581	5	3.1	290	12	US-10-040-655-7	Sequence 7, Appli
582	5	3.1	291	10	US-09-815-242-10544	Sequence 10544, A
583	5	3.1	293	9	US-09-992-598-309	Sequence 309, App
584	5	3.1	293	10	US-09-739-907-82	Sequence 82, Appl
585	5	3.1	293	10	US-09-739-907-97	Sequence 97, Appl
586	5	3.1	293	10	US-09-841-132-490	Sequence 490, App
587	5	3.1	293	10	US-09-989-722-309	Sequence 309, App
588	5	3.1	293	10	US-09-989-723-309	Sequence 309, App
589	5	3.1	293	10	US-09-989-279-309	Sequence 309, App
590	5	3.1	293	10	US-09-989-727-309	Sequence 309, App
591	5	3.1	293	10	US-09-989-731-309	Sequence 309, App
592	5	3.1	293	10	US-09-989-732-309	Sequence 309, App
593	5	3.1	293	10	US-09-991-073-309	Sequence 309, App
594	5	3.1	293	10	US-09-990-442-309	Sequence 309, App
595	5	3.1	293	10	US-09-991-163-309	Sequence 309, App
596	5	3.1	293	10	US-09-993-604-309	Sequence 309, App
597	5	3.1	293	10	US-09-990-456-309	Sequence 309, App
598	5	3.1	293	10	US-09-989-721-309	Sequence 309, App
599	5	3.1	294	10	US-09-767-037-3	Sequence 3, Appli
600	5	3.1	295	10	US-09-815-242-10104	Sequence 10104, A
601	5	3.1	298	9	US-09-860-670-138	Sequence 138, App
602	5	3.1	298	10	US-09-764-853-552	Sequence 552, App
603	5	3.1	298	10	US-09-764-877-1762	Sequence 1762, Ap
604	5	3.1	301	10	US-09-746-359A-19	Sequence 19, Appl
605	5	3.1	301	10	US-09-799-777-64	Sequence 64, Appl
606	5	3.1	301	10	US-09-764-877-2039	Sequence 2039, Ap
607	5	3.1	302	9	US-09-735-713A-4	Sequence 4, Appli
608	5	3.1	302	10	US-09-925-301-853	Sequence 853, App
609	5	3.1	303	10	US-09-733-569A-15	Sequence 15, Appl
610	5	3.1	303	10	US-09-917-340-4	Sequence 4, Appli
611	5	3.1	304	9	US-10-106-534-5	Sequence 5, Appli
612	5	3.1	305	10	US-09-886-055-299	Sequence 299, App
613	5	3.1	306	9	US-09-735-713A-2	Sequence 2, Appli
614	5	3.1	307	10	US-09-886-055-381	Sequence 381, App
615	5	3.1	308	9	US-09-895-913A-258	Sequence 258, App
616	5	3.1	308	10	US-09-999-162-2	Sequence 2, Appli
617	5	3.1	309	10	US-09-886-055-149	Sequence 149, App
618	5	3.1	309	10	US-09-886-055-471	Sequence 471, App
619	5	3.1	309	10	US-09-886-055-499	Sequence 499, App
620	5	3.1	310	9	US-10-042-194-1	Sequence 1, Appli
621	5	3.1	310	10	US-09-800-103-30	Sequence 30, Appl
622	5	3.1	310	10	US-09-886-055-485	Sequence 485, App
623	5	3.1	311	9	US-09-981-876-191	Sequence 191, App
624	5	3.1	311	10	US-09-864-761-46639	Sequence 46639, A
625	5	3.1	311	10	US-09-863-824-6	Sequence 6, Appli
626	5	3.1	312	10	US-09-886-055-265	Sequence 265, App
627	5	3.1	312	10	US-09-886-055-487	Sequence 487, App
628	5	3.1	313	10	US-09-764-864-1435	Sequence 1435, Ap
629	5	3.1	314	10	US-09-729-674-178	Sequence 178, App
630	5	3.1	314	10	US-09-815-242-5487	Sequence 5487, Ap
631	5	3.1	314	10	US-09-815-242-10584	Sequence 10584, A
632	5	3.1	314	10	US-09-771-209-80	Sequence 80, Appl
633	5	3.1	314	10	US-09-886-055-359	Sequence 359, App
634	5	3.1	314	10	US-09-771-730-41	Sequence 41, Appl
635	5	3.1	315	9	US-10-041-006A-9	Sequence 9, Appli
636	5	3.1	315	10	US-09-886-055-193	Sequence 193, App
637	5	3.1	315	10	US-09-886-055-435	Sequence 435, App
638	5	3.1	315	12	US-10-040-655-9	Sequence 9, Appli
639	5	3.1	316	10	US-09-886-055-203	Sequence 203, App
640	5	3.1	316	12	US-10-041-400A-9	Sequence 9, Appli
641	5	3.1	316	12	US-10-041-264A-9	Sequence 9, Appli
642	5	3.1	316	12	US-10-042-091A-9	Sequence 9, Appli
643	5	3.1	318	9	US-10-047-260-20	Sequence 20, Appl
644	5	3.1	318	10	US-09-759-143-920	Sequence 920, App
645	5	3.1	318	10	US-09-780-669-920	Sequence 920, App
646	5	3.1	318	10	US-09-822-827-920	Sequence 920, App
647	5	3.1	318	10	US-09-886-055-55	Sequence 55, Appl
648	5	3.1	318	10	US-09-886-055-91	Sequence 91, Appl
649	5	3.1	319	10	US-09-815-242-10267	Sequence 10267, A
650	5	3.1	319	10	US-09-767-037-4	Sequence 4, Appli
651	5	3.1	320	10	US-09-741-669-467	Sequence 467, App
652	5	3.1	321	10	US-09-771-730-43	Sequence 43, Appl
653	5	3.1	322	9	US-09-712-363-212	Sequence 212, App
654	5	3.1	322	10	US-09-788-626-14	Sequence 14, Appl
655	5	3.1	323	9	US-09-949-375A-2	Sequence 2, Appli
656	5	3.1	323	9	US-09-949-375A-4	Sequence 4, Appli
657	5	3.1	323	9	US-09-949-375A-6	Sequence 6, Appli
658	5	3.1	323	10	US-09-815-242-11034	Sequence 11034, A
659	5	3.1	324	10	US-09-815-242-5117	Sequence 5117, Ap
660	5	3.1	325	9	US-10-033-245-11	Sequence 11, Appl
661	5	3.1	325	9	US-10-033-223-11	Sequence 11, Appl
662	5	3.1	325	12	US-10-033-246-11	Sequence 11, Appl
663	5	3.1	325	12	US-10-033-301-11	Sequence 11, Appl
664	5	3.1	325	12	US-10-033-326-11	Sequence 11, Appl
665	5	3.1	326	10	US-09-371-307-65	Sequence 65, Appl
666	5	3.1	326	10	US-09-815-242-12589	Sequence 12589, A
667	5	3.1	327	10	US-09-815-242-5652	Sequence 5652, Ap
668	5	3.1	327	10	US-09-815-242-11970	Sequence 11970, A
669	5	3.1	327	10	US-09-769-066-17	Sequence 17, Appl
670	5	3.1	327	10	US-09-769-066-18	Sequence 18, Appl
671	5	3.1	328	9	US-09-978-295A-358	Sequence 358, App
672	5	3.1	328	9	US-09-978-697-358	Sequence 358, App
673	5	3.1	328	10	US-09-864-761-35526	Sequence 35526, A
674	5	3.1	328	10	US-09-815-242-12436	Sequence 12436, A
675	5	3.1	329	8	US-08-945-038-4	Sequence 4, Appli
676	5	3.1	329	10	US-09-815-242-11142	Sequence 11142, A

677	5	3.1	330	9	US-09-949-375A-10	Sequence 10, Appl
678	5	3.1	330	10	US-09-794-591-2	Sequence 2, Appli
679	5	3.1	330	10	US-09-925-302-453	Sequence 453, App
680	5	3.1	331	10	US-09-401-636-1	Sequence 1, Appli
681	5	3.1	331	10	US-09-815-242-5912	Sequence 5912, Ap
682	5	3.1	331	10	US-09-815-242-12872	Sequence 12872, A
683	5	3.1	331	10	US-09-815-242-13146	Sequence 13146, A
684	5	3.1	332	9	US-09-764-868-1013	Sequence 1013, Ap
685	5	3.1	332	9	US-09-949-375A-23	Sequence 23, Appl
686	5	3.1	332	9	US-09-949-375A-25	Sequence 25, Appl
687	5	3.1	332	9	US-09-949-375A-27	Sequence 27, Appl
688	5	3.1	332	10	US-09-925-302-484	Sequence 484, App
689	5	3.1	332	10	US-09-925-300-1064	Sequence 1064, Ap
690	5	3.1	334	10	US-09-764-864-1497	Sequence 1497, Ap
691	5	3.1	335	10	US-09-815-242-13433	Sequence 13433, A
692	5	3.1	335	10	US-09-815-242-13658	Sequence 13658, A
693	5	3.1	335	10	US-09-821-687-11	Sequence 11, Appl
694	5	3.1	336	9	US-09-949-375A-8	Sequence 8, Appli
695	5	3.1	336	10	US-09-789-919-52	Sequence 52, Appl
696	5	3.1	337	10	US-09-946-143-2	Sequence 2, Appli
697	5	3.1	338	10	US-09-815-242-5317	Sequence 5317, Ap
698	5	3.1	339	10	US-09-815-242-5196	Sequence 5196, Ap
699	5	3.1	339	10	US-09-943-671-2	Sequence 2, Appli
700	5	3.1	340	10	US-09-159-469-54	Sequence 54, Appl
701	5	3.1	340	10	US-09-798-042-54	Sequence 54, Appl
702	5	3.1	341	10	US-09-401-636-6	Sequence 6, Appli
703	5	3.1	341	10	US-09-401-636-11	Sequence 11, Appl
704	5	3.1	341	10	US-09-948-649-2	Sequence 2, Appli
705	5	3.1	342	10	US-09-401-636-8	Sequence 8, Appli
706	5	3.1	342	10	US-09-849-980B-1	Sequence 1, Appli
707	5	3.1	343	9	US-09-949-375A-20	Sequence 20, Appl
708	5	3.1	343	9	US-09-949-375A-22	Sequence 22, Appl
709	5	3.1	343	10	US-09-864-761-46837	Sequence 46837, A
710	5	3.1	343	10	US-09-925-301-981	Sequence 981, App
711	5	3.1	343	10	US-09-815-242-10474	Sequence 10474, A
712	5	3.1	344	9	US-09-712-363-229	Sequence 229, App
713	5	3.1	344	10	US-09-815-242-12624	Sequence 12624, A
714	5	3.1	344	10	US-09-898-570-18	Sequence 18, Appl
715	5	3.1	345	10	US-09-401-636-10	Sequence 10, Appl
716	5	3.1	345	10	US-09-815-242-5354	Sequence 5354, Ap
717	5	3.1	346	10	US-09-864-761-47650	Sequence 47650, A
718	5	3.1	347	10	US-09-815-242-11680	Sequence 11680, A
719	5	3.1	347	10	US-09-789-561-184	Sequence 184, App
720	5	3.1	348	10	US-09-318-271-10	Sequence 10, Appl
721	5	3.1	348	10	US-09-864-761-37014	Sequence 37014, A
722	5	3.1	348	10	US-09-886-055-433	Sequence 433, App
723	5	3.1	349	10	US-09-815-242-10194	Sequence 10194, A
724	5	3.1	349	10	US-09-815-242-13751	Sequence 13751, A
725	5	3.1	351	9	US-09-829-378-5	Sequence 5, Appli
726	5	3.1	354	10	US-09-886-055-331	Sequence 331, App
727	5	3.1	355	10	US-09-874-132-3	Sequence 3, Appli
728	5	3.1	356	10	US-09-925-637-62	Sequence 62, Appl
729	5	3.1	356	10	US-09-735-367B-13	Sequence 13, Appl
730	5	3.1	357	12	US-10-093-045-8	Sequence 8, Appli
731	5	3.1	357	12	US-10-093-045-9	Sequence 9, Appli
732	5	3.1	358	10	US-09-287-849-8	Sequence 8, Appli
733	5	3.1	358	10	US-09-784-508-4	Sequence 4, Appli
734	5	3.1	359	9	US-09-972-758-2	Sequence 2, Appli
735	5	3.1	359	9	US-09-828-523A-14	Sequence 14, Appl
736	5	3.1	359	10	US-09-925-301-1127	Sequence 1127, Ap
737	5	3.1	359	10	US-09-815-242-10446	Sequence 10446, A
738	5	3.1	359	10	US-09-815-242-14026	Sequence 14026, A
739	5	3.1	359	10	US-09-888-876-2	Sequence 2, Appli
740	5	3.1	359	10	US-09-745-763-34	Sequence 34, Appl
741	5	3.1	359	10	US-09-951-622-13	Sequence 13, Appl
742	5	3.1	359	12	US-10-093-045-1	Sequence 1, Appli
743	5	3.1	360	10	US-09-815-242-11254	Sequence 11254, A
744	5	3.1	362	10	US-09-925-301-1000	Sequence 1000, Ap
745	5	3.1	362	12	US-10-093-045-2	Sequence 2, Appli
746	5	3.1	363	10	US-09-925-301-1271	Sequence 1271, Ap
747	5	3.1	363	10	US-09-815-242-11979	Sequence 11979, A
748	5	3.1	363	10	US-09-764-864-1093	Sequence 1093, Ap
749	5	3.1	363	10	US-09-764-864-1513	Sequence 1513, Ap
750	5	3.1	364	10	US-09-815-242-13875	Sequence 13875, A
751	5	3.1	364	10	US-09-764-903-57	Sequence 57, Appl
752	5	3.1	364	10	US-09-756-983-22	Sequence 22, Appl
753	5	3.1	366	10	US-09-815-242-11166	Sequence 11166, A
754	5	3.1	366	10	US-09-758-008-3	Sequence 3, Appli
755	5	3.1	367	10	US-09-815-242-10676	Sequence 10676, A
756	5	3.1	369	9	US-09-712-363-170	Sequence 170, App
757	5	3.1	370	9	US-09-828-523A-74	Sequence 74, Appl
758	5	3.1	370	10	US-09-205-658-107	Sequence 107, App
759	5	3.1	370	10	US-09-844-353A-107	Sequence 107, App
760	5	3.1	372	10	US-09-739-907-179	Sequence 179, App
761	5	3.1	372	10	US-09-815-242-5709	Sequence 5709, Ap
762	5	3.1	375	10	US-09-853-386-135	Sequence 135, App
763	5	3.1	375	10	US-09-853-386-141	Sequence 141, App
764	5	3.1	375	10	US-09-853-386-143	Sequence 143, App
765	5	3.1	375	10	US-09-828-313-34	Sequence 34, Appl
766	5	3.1	377	9	US-09-916-494A-14	Sequence 14, Appl
767	5	3.1	377	9	US-10-091-628-2	Sequence 2, Appli
768	5	3.1	377	10	US-09-881-752A-92	Sequence 92, Appl
769	5	3.1	378	9	US-09-970-989-17	Sequence 17, Appl
770	5	3.1	378	9	US-09-992-598-156	Sequence 156, App
771	5	3.1	378	10	US-09-816-664-7	Sequence 7, Appli
772	5	3.1	378	10	US-09-815-242-13446	Sequence 13446, A
773	5	3.1	378	10	US-09-989-722-156	Sequence 156, App
774	5	3.1	378	10	US-09-989-723-156	Sequence 156, App
775	5	3.1	378	10	US-09-989-279-156	Sequence 156, App
776	5	3.1	378	10	US-09-989-727-156	Sequence 156, App
777	5	3.1	378	10	US-09-989-731-156	Sequence 156, App
778	5	3.1	378	10	US-09-989-732-156	Sequence 156, App
779	5	3.1	378	10	US-09-991-073-156	Sequence 156, App
780	5	3.1	378	10	US-09-990-442-156	Sequence 156, App
781	5	3.1	378	10	US-09-991-163-156	Sequence 156, App
782	5	3.1	378	10	US-09-993-604-156	Sequence 156, App
783	5	3.1	378	10	US-09-990-456-156	Sequence 156, App
784	5	3.1	378	10	US-09-989-721-156	Sequence 156, App
785	5	3.1	379	9	US-09-860-846-16	Sequence 16, Appl
786	5	3.1	379	10	US-09-861-289-16	Sequence 16, Appl
787	5	3.1	380	9	US-10-020-674-2	Sequence 2, Appli
788	5	3.1	380	10	US-09-823-240-10	Sequence 10, Appl
789	5	3.1	381	10	US-09-205-658-106	Sequence 106, App
790	5	3.1	381	10	US-09-844-353A-106	Sequence 106, App
791	5	3.1	382	9	US-09-975-139-3	Sequence 3, Appli
792	5	3.1	382	10	US-09-825-414-30	Sequence 30, Appl
793	5	3.1	382	10	US-09-919-831-3	Sequence 3, Appli
794	5	3.1	383	9	US-09-764-868-896	Sequence 896, App
795	5	3.1	383	9	US-10-047-542-28	Sequence 28, Appl
796	5	3.1	383	9	US-10-047-542-32	Sequence 32, Appl
797	5	3.1	383	10	US-09-147-346-2	Sequence 2, Appli
798	5	3.1	384	9	US-09-852-797-118	Sequence 118, App
799	5	3.1	384	10	US-09-758-498-3	Sequence 3, Appli
800	5	3.1	384	10	US-09-853-161-118	Sequence 118, App
801	5	3.1	384	10	US-09-852-659A-118	Sequence 118, App
802	5	3.1	385	9	US-09-963-234-10	Sequence 10, Appl
803	5	3.1	385	10	US-09-815-242-5713	Sequence 5713, Ap
804	5	3.1	386	10	US-09-808-387-6	Sequence 6, Appli
805	5	3.1	387	10	US-09-794-715A-4	Sequence 4, Appli
806	5	3.1	388	10	US-09-767-041-33	Sequence 33, Appl
807	5	3.1	388	10	US-09-815-242-11458	Sequence 11458, A
808	5	3.1	388	10	US-09-815-242-11623	Sequence 11623, A
809	5	3.1	388	10	US-09-881-752A-350	Sequence 350, App
810	5	3.1	388	10	US-09-794-715A-2	Sequence 2, Appli
811	5	3.1	388	10	US-09-924-256A-94	Sequence 94, Appl
812	5	3.1	388	12	US-10-052-586-336	Sequence 336, App
813	5	3.1	390	10	US-09-925-297-531	Sequence 531, App
814	5	3.1	391	10	US-09-815-242-5655	Sequence 5655, Ap
815	5	3.1	391	10	US-09-815-242-12953	Sequence 12953, A
816	5	3.1	391	10	US-09-815-242-13119	Sequence 13119, A
817	5	3.1	392	9	US-09-935-124A-2	Sequence 2, Appli
818	5	3.1	392	10	US-09-925-637-8	Sequence 8, Appli
819	5	3.1	393	10	US-09-815-242-12395	Sequence 12395, A
820	5	3.1	393	10	US-09-745-763-19	Sequence 19, Appl
821	5	3.1	394	10	US-09-853-386-33	Sequence 33, Appl
822	5	3.1	394	10	US-09-815-242-12267	Sequence 12267, A

823	5	3.1	395	10	US-09-952-432A-19	Sequence 19, Appl
824	5	3.1	395	10	US-09-952-432A-21	Sequence 21, Appl
825	5	3.1	397	10	US-09-925-301-1138	Sequence 1138, Ap
826	5	3.1	398	10	US-09-815-242-10390	Sequence 10390, A
827	5	3.1	399	10	US-09-764-870-409	Sequence 409, App
828	5	3.1	400	10	US-09-922-217-1115	Sequence 1115, Ap
829	5	3.1	401	10	US-09-853-386-127	Sequence 127, App
830	5	3.1	401	10	US-09-853-386-139	Sequence 139, App
831	5	3.1	401	10	US-09-815-242-13828	Sequence 13828, A
832	5	3.1	402	10	US-09-838-564A-2	Sequence 2, Appli
833	5	3.1	402	10	US-09-838-564A-5	Sequence 5, Appli
834	5	3.1	402	10	US-09-933-561-16	Sequence 16, Appl
835	5	3.1	404	9	US-09-712-363-257	Sequence 257, App
836	5	3.1	404	10	US-09-815-242-11242	Sequence 11242, A
837	5	3.1	404	10	US-09-815-242-14052	Sequence 14052, A
838	5	3.1	404	10	US-09-764-864-1037	Sequence 1037, Ap
839	5	3.1	405	10	US-09-864-761-38102	Sequence 38102, A
840	5	3.1	405	12	US-10-060-333-2	Sequence 2, Appli
841	5	3.1	406	10	US-09-851-588-6	Sequence 6, Appli
842	5	3.1	406	10	US-09-898-570-20	Sequence 20, Appl
843	5	3.1	408	9	US-09-712-363-268	Sequence 268, App
844	5	3.1	409	9	US-09-764-868-646	Sequence 646, App
845	5	3.1	409	10	US-09-790-264-42	Sequence 42, Appl
846	5	3.1	409	10	US-09-815-242-14050	Sequence 14050, A
847	5	3.1	410	10	US-09-804-551B-18	Sequence 18, Appl
848	5	3.1	410	10	US-09-963-896-2	Sequence 2, Appli
849	5	3.1	411	10	US-09-764-870-351	Sequence 351, App
850	5	3.1	411	10	US-09-881-401-4	Sequence 4, Appli
851	5	3.1	411	10	US-09-881-401-8	Sequence 8, Appli
852	5	3.1	412	10	US-09-925-300-1669	Sequence 1669, Ap
853	5	3.1	413	10	US-09-925-299-977	Sequence 977, App
854	5	3.1	415	10	US-09-853-386-128	Sequence 128, App
855	5	3.1	415	10	US-09-853-386-130	Sequence 130, App
856	5	3.1	415	10	US-09-853-386-131	Sequence 131, App
857	5	3.1	415	10	US-09-853-386-133	Sequence 133, App
858	5	3.1	415	10	US-09-853-386-134	Sequence 134, App
859	5	3.1	415	10	US-09-853-386-137	Sequence 137, App
860	5	3.1	415	10	US-09-853-386-138	Sequence 138, App
861	5	3.1	415	10	US-09-925-301-1360	Sequence 1360, Ap
862	5	3.1	415	10	US-09-191-724-2	Sequence 2, Appli
863	5	3.1	415	10	US-09-191-724-6	Sequence 6, Appli
864	5	3.1	415	10	US-09-191-724-13	Sequence 13, Appl
865	5	3.1	417	10	US-09-781-867-2	Sequence 2, Appli
866	5	3.1	418	10	US-09-808-483-18	Sequence 18, Appl
867	5	3.1	418	10	US-09-925-302-513	Sequence 513, App
868	5	3.1	419	10	US-09-925-300-1506	Sequence 1506, Ap
869	5	3.1	420	10	US-09-853-386-129	Sequence 129, App
870	5	3.1	421	9	US-10-046-961-5	Sequence 5, Appli
871	5	3.1	421	9	US-09-949-375A-28	Sequence 28, Appl
872	5	3.1	421	10	US-09-420-785A-4	Sequence 4, Appli
873	5	3.1	421	10	US-09-901-252-15	Sequence 15, Appl
874	5	3.1	421	10	US-09-815-242-5303	Sequence 5303, Ap
875	5	3.1	421	10	US-09-815-242-12388	Sequence 12388, A
876	5	3.1	421	10	US-09-815-242-12820	Sequence 12820, A
877	5	3.1	421	10	US-09-793-306-163	Sequence 163, App
878	5	3.1	422	10	US-09-815-242-5208	Sequence 5208, Ap
879	5	3.1	422	10	US-09-815-242-11329	Sequence 11329, A
880	5	3.1	422	10	US-09-815-242-11825	Sequence 11825, A
881	5	3.1	422	10	US-09-924-338-2	Sequence 2, Appli
882	5	3.1	422	10	US-09-779-307-18	Sequence 18, Appl
883	5	3.1	426	10	US-09-815-242-13849	Sequence 13849, A
884	5	3.1	427	9	US-09-850-982B-2	Sequence 2, Appli
885	5	3.1	427	9	US-09-975-253-2	Sequence 2, Appli
886	5	3.1	427	10	US-09-764-864-1086	Sequence 1086, Ap
887	5	3.1	428	9	US-10-047-542-60	Sequence 60, Appl
888	5	3.1	428	9	US-09-949-375A-1	Sequence 1, Appli
889	5	3.1	428	10	US-09-916-230-1	Sequence 1, Appli
890	5	3.1	429	9	US-10-047-542-30	Sequence 30, Appl
891	5	3.1	429	9	US-10-157-855-24	Sequence 24, Appl
892	5	3.1	430	10	US-09-864-761-34568	Sequence 34568, A
893	5	3.1	430	10	US-09-853-386-140	Sequence 140, App
894	5	3.1	430	10	US-09-815-242-11315	Sequence 11315, A
895	5	3.1	430	12	US-10-026-001-1	Sequence 1, Appli
896	5	3.1	431	10	US-09-191-724-10	Sequence 10, Appl
897	5	3.1	431	10	US-09-881-401-2	Sequence 2, Appli
898	5	3.1	432	9	US-10-157-855-25	Sequence 25, Appl
899	5	3.1	432	9	US-09-949-375A-19	Sequence 19, Appl
900	5	3.1	432	10	US-09-809-517A-25	Sequence 25, Appl
901	5	3.1	432	10	US-09-864-761-43246	Sequence 43246, A
902	5	3.1	432	10	US-09-919-172-9	Sequence 9, Appli
903	5	3.1	432	12	US-10-006-867-112	Sequence 112, App
904	5	3.1	432	12	US-10-052-586-330	Sequence 330, App
905	5	3.1	433	9	US-09-866-570A-66	Sequence 66, Appl
906	5	3.1	433	10	US-09-866-572A-66	Sequence 66, Appl
907	5	3.1	434	10	US-09-809-517A-23	Sequence 23, Appl
908	5	3.1	434	10	US-09-809-517A-26	Sequence 26, Appl
909	5	3.1	434	10	US-09-815-242-10133	Sequence 10133, A
910	5	3.1	434	10	US-09-771-161A-92	Sequence 92, Appl
911	5	3.1	435	9	US-09-895-913A-146	Sequence 146, App
912	5	3.1	435	10	US-09-801-368-150	Sequence 150, App
913	5	3.1	436	9	US-09-866-570A-62	Sequence 62, Appl
914	5	3.1	436	10	US-09-240-675-2	Sequence 2, Appli
915	5	3.1	436	10	US-09-922-501-8	Sequence 8, Appli
916	5	3.1	436	10	US-09-866-572A-62	Sequence 62, Appl
917	5	3.1	437	10	US-09-790-264-30	Sequence 30, Appl
918	5	3.1	437	10	US-09-851-588-8	Sequence 8, Appli
919	5	3.1	440	10	US-09-919-497-80	Sequence 80, Appl
920	5	3.1	440	10	US-09-847-081B-2	Sequence 2, Appli
921	5	3.1	441	9	US-09-949-375A-7	Sequence 7, Appli
922	5	3.1	441	10	US-09-924-338-4	Sequence 4, Appli
923	5	3.1	444	9	US-09-464-099A-53	Sequence 53, Appl
924	5	3.1	444	9	US-09-464-099A-54	Sequence 54, Appl
925	5	3.1	444	9	US-09-464-099A-55	Sequence 55, Appl
926	5	3.1	444	9	US-09-464-099A-56	Sequence 56, Appl
927	5	3.1	444	10	US-09-861-696-53	Sequence 53, Appl
928	5	3.1	444	10	US-09-861-696-54	Sequence 54, Appl
929	5	3.1	444	10	US-09-861-696-55	Sequence 55, Appl
930	5	3.1	444	10	US-09-861-696-56	Sequence 56, Appl
931	5	3.1	444	10	US-09-853-386-132	Sequence 132, App
932	5	3.1	444	10	US-09-853-386-136	Sequence 136, App
933	5	3.1	444	10	US-09-191-724-15	Sequence 15, Appl
934	5	3.1	444	10	US-09-840-795-8	Sequence 8, Appli
935	5	3.1	444	10	US-09-747-835A-49	Sequence 49, Appl
936	5	3.1	445	9	US-10-108-605-263	Sequence 263, App
937	5	3.1	446	10	US-09-951-780-2	Sequence 2, Appli
938	5	3.1	446	10	US-09-903-456-30	Sequence 30, Appl
939	5	3.1	447	10	US-09-888-615-109	Sequence 109, App
940	5	3.1	450	10	US-09-790-264-34	Sequence 34, Appl
941	5	3.1	451	9	US-10-074-279-6	Sequence 6, Appli
942	5	3.1	452	1	US-08-841-636A-35	Sequence 35, Appl
943	5	3.1	454	10	US-09-746-491-28	Sequence 28, Appl
944	5	3.1	456	10	US-09-861-400-6	Sequence 6, Appli
945	5	3.1	457	9	US-09-712-363-198	Sequence 198, App
946	5	3.1	460	10	US-09-989-861-18	Sequence 18, Appl
947	5	3.1	461	9	US-10-046-961-6	Sequence 6, Appli
948	5	3.1	461	9	US-09-712-363-285	Sequence 285, App
949	5	3.1	462	10	US-09-815-242-10991	Sequence 10991, A
950	5	3.1	462	12	US-10-052-586-298	Sequence 298, App
951	5	3.1	463	9	US-09-895-913A-348	Sequence 348, App
952	5	3.1	463	10	US-09-858-075A-2	Sequence 2, Appli
953	5	3.1	463	10	US-09-815-242-11325	Sequence 11325, A
954	5	3.1	463	12	US-10-083-452-10	Sequence 10, Appl
955	5	3.1	464	10	US-09-863-824-2	Sequence 2, Appli
956	5	3.1	467	10	US-09-841-880-4	Sequence 4, Appli
957	5	3.1	468	10	US-09-860-232A-8	Sequence 8, Appli
958	5	3.1	470	10	US-09-815-242-12078	Sequence 12078, A
959	5	3.1	471	10	US-09-881-752A-18	Sequence 18, Appl
960	5	3.1	474	12	US-10-007-693-99	Sequence 99, Appl
961	5	3.1	476	10	US-09-779-307-16	Sequence 16, Appl
962	5	3.1	476	12	US-10-042-991-2	Sequence 2, Appli
963	5	3.1	477	10	US-09-779-307-4	Sequence 4, Appli
964	5	3.1	479	9	US-09-813-398-26	Sequence 26, Appl
965	5	3.1	479	10	US-09-925-300-1515	Sequence 1515, Ap
966	5	3.1	480	10	US-09-790-264-32	Sequence 32, Appl
967	5	3.1	480	10	US-09-802-472B-4	Sequence 4, Appli
968	5	3.1	480	12	US-10-042-991-3	Sequence 3, Appli

969	5	3.1	481	12	US-10-105-929-8	Sequence 8, Appli
970	5	3.1	482	10	US-09-779-307-17	Sequence 17, Appl
971	5	3.1	482	10	US-09-861-400-5	Sequence 5, Appli
972	5	3.1	483	9	US-09-974-298-41	Sequence 41, Appl
973	5	3.1	483	9	US-09-981-353-158	Sequence 158, App
974	5	3.1	483	10	US-09-739-254-154	Sequence 154, App
975	5	3.1	483	10	US-09-732-618-10	Sequence 10, Appl
976	5	3.1	483	10	US-09-904-615-154	Sequence 154, App
977	5	3.1	483	10	US-09-805-458A-6	Sequence 6, Appli
978	5	3.1	483	10	US-09-919-497-79	Sequence 79, Appl
979	5	3.1	483	12	US-10-042-991-4	Sequence 4, Appli
980	5	3.1	484	10	US-09-790-264-36	Sequence 36, Appl
981	5	3.1	484	10	US-09-815-242-11054	Sequence 11054, A
982	5	3.1	484	12	US-10-006-867-78	Sequence 78, Appl
983	5	3.1	484	12	US-10-052-586-280	Sequence 280, App
984	5	3.1	485	9	US-09-860-846-10	Sequence 10, Appl
985	5	3.1	485	10	US-09-861-289-10	Sequence 10, Appl
986	5	3.1	487	10	US-09-815-242-5183	Sequence 5183, Ap
987	5	3.1	487	10	US-09-825-414-66	Sequence 66, Appl
988	5	3.1	488	10	US-09-884-260A-38	Sequence 38, Appl
989	5	3.1	488	12	US-10-042-991-6	Sequence 6, Appli
990	5	3.1	491	10	US-09-881-752A-128	Sequence 128, App
991	5	3.1	493	10	US-09-815-242-12052	Sequence 12052, A
992	5	3.1	494	10	US-09-815-242-10507	Sequence 10507, A
993	5	3.1	494	12	US-10-143-002-4	Sequence 4, Appli
994	5	3.1	496	9	US-09-963-234-2	Sequence 2, Appli
995	5	3.1	496	10	US-09-974-592-10	Sequence 10, Appl
996	5	3.1	497	10	US-09-734-676-2	Sequence 2, Appli
997	5	3.1	498	10	US-09-925-302-475	Sequence 475, App
998	5	3.1	498	10	US-09-764-864-1122	Sequence 1122, Ap
999	5	3.1	499	10	US-09-734-676-4	Sequence 4, Appli
1000	5	3.1	500	9	US-10-029-009-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-09-900-575-29
; Sequence 29, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 29
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-29

Query Match 100.0%; Score 161; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 4.3e-148;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGGLVIKAGSLIAVLILRQTNNYNSDDFQFV 145

Qy     121 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||
Db     146 WNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186
```

RESULT 2
US-09-900-575-34
; Sequence 34, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-34

Query Match 57.1%; Score 92; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.7e-81;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||||
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117
```

RESULT 3
US-09-900-575-36
; Sequence 36, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
; LENGTH: 279
; TYPE: PRT
; ORGANISM: E. coli
US-09-900-575-36

Query Match 57.1%; Score 92; DB 10; Length 279;
Best Local Similarity 100.0%; Pred. No. 1.7e-81;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy      61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||||
```

```
Db      86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 4
US-09-900-575-44
; Sequence 44, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-44

  Query Match           57.1%;  Score 92;  DB 10;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-81;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 85

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db     86 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 5
US-09-912-020-367
; Sequence 367, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Trawick, John
;  APPLICANT:  Forsyth, R. Allyn
;  APPLICANT:  Froelich, Jamie M.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001DV1
;  CURRENT APPLICATION NUMBER: US/09/912,020
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 367
;   LENGTH: 300
;   TYPE: PRT
;   ORGANISM: E. Coli
US-09-912-020-367

  Query Match           57.1%;  Score 92;  DB 10;  Length 300;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-81;
  Matches   92;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     47 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFP 106

Qy     61 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||||
Db    107 TTSETPRVVYNSRTDKPWPVALYLTPVSSAGG 138


RESULT 6
US-09-900-575-27
; Sequence 27, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-27

  Query Match           55.9%;  Score 90;  DB 10;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-79;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy     63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
          ||||||||||||||||||||||||||||||
Db     88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 7
US-09-900-575-28
; Sequence 28, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 28
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-28

  Query Match           55.9%;  Score 90;  DB 10;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-79;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 8
US-09-900-575-32
; Sequence 32, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-32

  Query Match             55.9%;  Score 90;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-79;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 9
US-09-900-575-39
; Sequence 39, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 39
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-39

  Query Match             55.9%;  Score 90;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-79;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 10
US-09-900-575-43
; Sequence 43, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 43
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-43

  Query Match             55.9%;  Score 90;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-79;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 11
US-09-900-575-55
; Sequence 55, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 55
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Consensus sequence of FimH proteins for SEQ ID NO: 23 to 45
US-09-900-575-55

  Query Match             55.9%;  Score 90;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-79;
  Matches   90;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 62
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 VNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTT 87

Qy      63 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 92
           ||||||||||||||||||||||||||||||
Db      88 SETPRVVYNSRTDKPWPVALYLTPVSSAGG 117


RESULT 12
US-09-900-575-30
; Sequence 30, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
;   LENGTH: 280
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-30

  Query Match             46.6%;  Score 75;  DB 10;  Length 280;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-65;
  Matches   75;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      18 FCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKP 77
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43 FCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKP 102

Qy      78 WPVALYLTPVSSAGG 92
           ||||||||||||||||
Db     103 WPVALYLTPVSSAGG 117


RESULT 13
US-09-900-575-25
; Sequence 25, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-25

  Query Match             41.6%;  Score 67;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-57;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 14
US-09-900-575-26
; Sequence 26, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-26

  Query Match             41.6%;  Score 67;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-57;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 15
US-09-900-575-35
; Sequence 35, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-35

  Query Match             41.6%;  Score 67;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-57;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179
```

```
Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 16
US-09-900-575-37
; Sequence 37, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 37
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-37

  Query Match           41.6%;  Score 67;  DB 10;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-57;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 17
US-09-900-575-40
; Sequence 40, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-40

  Query Match           41.6%;  Score 67;  DB 10;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-57;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 18
US-09-900-575-41
; Sequence 41, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 41
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-41

  Query Match           41.6%;  Score 67;  DB 10;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-57;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 19
US-09-900-575-42
; Sequence 42, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 42
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-42

  Query Match           41.6%;  Score 67;  DB 10;  Length 279;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-57;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
```

```
              |||||||
Db    180 PIPLTVY 186


RESULT 20
US-09-900-575-45
; Sequence 45, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 45
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-45

  Query Match             41.6%;  Score 67;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-57;
  Matches   67;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 154
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSV 179

Qy     155 PIPLTVY 161
           |||||||
Db     180 PIPLTVY 186


RESULT 21
US-09-900-575-38
; Sequence 38, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-38

  Query Match             37.3%;  Score 60;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-50;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     102 AVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 161
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     127 AVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVY 186


RESULT 22
US-09-900-575-33
; Sequence 33, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-33

  Query Match             34.8%;  Score 56;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-47;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 150
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDY 175


RESULT 23
US-09-900-575-23
; Sequence 23, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-23

  Query Match             29.8%;  Score 48;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-39;
  Matches   48;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG 48
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db      26 PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG 73


RESULT 24
US-09-900-575-24
; Sequence 24, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
```

```
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-24

  Query Match             26.7%;  Score 43;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-34;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCD 137
           |||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCD 162


RESULT 25
US-09-900-575-31
; Sequence 31, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
;   LENGTH: 279
;   TYPE: PRT
;   ORGANISM: E. coli
US-09-900-575-31

  Query Match             26.7%;  Score 43;  DB 10;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-34;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      95 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCD 137
           |||||||||||||||||||||||||||||||||||||||||||||
Db     120 IKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCD 162


RESULT 26
US-10-014-882-2
; Sequence 2, Application US/10014882
; Patent No. US20020107384A1
; GENERAL INFORMATION:
;  APPLICANT: Hu, Yi
;  APPLICANT:  Kieke, James
;  APPLICANT:  Donoho, Gregory
;  TITLE OF INVENTION: No. US20020107384A1el Human Kinase and Polynucleotides Encoding t
;  FILE REFERENCE: LEX-0279-USA
;  CURRENT APPLICATION NUMBER: US/10/014,882
;  CURRENT FILING DATE:  2001-12-11
;  PRIOR APPLICATION NUMBER: US 60/254,744
;  PRIOR FILING DATE: 2000-12-11
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1036
;   TYPE: PRT
;   ORGANISM: homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(1036)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-10-014-882-2

  Query Match             5.0%;  Score 8;  DB 12;  Length 1036;
  Best Local Similarity   100.0%;  Pred. No. 6.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 TPVSSAGG 92
           ||||||||
Db      12 TPVSSAGG 19


RESULT 27
US-09-901-436A-11
; Sequence 11, Application US/09901436A
; Patent No. US20020098515A1
; GENERAL INFORMATION:
;  APPLICANT: Ni et al.
;  TITLE OF INVENTION: Cytostatin I
;  FILE REFERENCE: PF175D4
;  CURRENT APPLICATION NUMBER: US/09/901,436A
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: 09/361,737
;  PRIOR FILING DATE: 1999-07-28
;  PRIOR APPLICATION NUMBER: 09/023,073
;  PRIOR FILING DATE: 1998-02-13
;  PRIOR APPLICATION NUMBER: 08/470,298
;  PRIOR FILING DATE: 1995-06-06
;  PRIOR APPLICATION NUMBER: 08/409,731
;  PRIOR FILING DATE: 1995-03-24
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
;   LENGTH: 132
;   TYPE: PRT
;   ORGANISM: human
US-09-901-436A-11

  Query Match             4.3%;  Score 7;  DB 10;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 9;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 VTLQRGS 37
           |||||||
Db      85 VTLQRGS 91


RESULT 28
US-09-912-020-282
; Sequence 282, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
;  APPLICANT: Zyskind, Judith
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Trawick, John
;  APPLICANT:  Forsyth, R. Allyn
;  APPLICANT:  Froelich, Jamie M.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;  TITLE OF INVENTION:  ESCHERICHIA COLI
;  FILE REFERENCE: ELITRA.001DV1
;  CURRENT APPLICATION NUMBER: US/09/912,020
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: 09/492,709
;  PRIOR FILING DATE: 2000-01-27
;  PRIOR APPLICATION NUMBER: 60/117,405
;  PRIOR FILING DATE: 1999-01-27
```

```
;  NUMBER OF SEQ ID NOS: 485
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 282
;   LENGTH: 1040
;   TYPE: PRT
;   ORGANISM: E. Coli
US-09-912-020-282

  Query Match             4.3%;  Score 7;  DB 10;  Length 1040;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 SLIAVLI 105
           |||||||
Db     450 SLIAVLI 456


RESULT 29
US-09-815-242-10203
; Sequence 10203, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10203
;   LENGTH: 1040
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-815-242-10203

  Query Match             4.3%;  Score 7;  DB 10;  Length 1040;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 SLIAVLI 105
           |||||||
Db     450 SLIAVLI 456


RESULT 30
US-09-864-761-37358
; Sequence 37358, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37358
;   LENGTH: 40
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC009721.9
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.9
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.2
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.5
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
US-09-864-761-37358

  Query Match             3.7%;  Score 6;  DB 10;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      33 LQRGSA 38
           ||||||
Db       7 LQRGSA 12


RESULT 31
```

```
US-09-726-643-65
; Sequence 65, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
;  APPLICANT:  Ruben et al.
;  TITLE OF INVENTION:  26 Human secreted proteins
;  FILE REFERENCE:  PZ040P1
;  CURRENT APPLICATION NUMBER: US/09/726,643
;  CURRENT FILING DATE:  2000-12-01
;  PRIOR APPLICATION NUMBER:  PCT/US00/15187
;  PRIOR FILING DATE:  2000-06-02
;  PRIOR APPLICATION NUMBER:  60/137,725
;  PRIOR FILING DATE:  1999-06-07
;  NUMBER OF SEQ ID NOS:  190
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 65
;   LENGTH: 46
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-726-643-65

  Query Match             3.7%;  Score 6;  DB 10;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     143 VTVTLP 148
           ||||||
Db       3 VTVTLP 8


RESULT 32
US-09-864-761-34310
; Sequence 34310, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34310
;   LENGTH: 54
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC009238.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.93
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.73
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
;   OTHER INFORMATION: SWISSPROT HIT: Q59050, EVALUE 6.00e-07
;   OTHER INFORMATION: EST_HUMAN HIT: AI884989.1, EVALUE 3.00e-26
US-09-864-761-34310

  Query Match             3.7%;  Score 6;  DB 10;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     137 DVSARD 142
           ||||||
Db      16 DVSARD 21


RESULT 33
US-09-864-761-35614
; Sequence 35614, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
```

```
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35614
;   LENGTH: 62
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO Z98044.10
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.6
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
;   OTHER INFORMATION: SWISSPROT HIT: O95479, EVALUE 7.00e-30
;   OTHER INFORMATION: EST_HUMAN HIT: AW503925.1, EVALUE 6.00e-29
US-09-864-761-35614

  Query Match             3.7%;  Score 6;  DB 10;  Length 62;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      33 LQRGSA 38
           ||||||
Db      27 LQRGSA 32


RESULT 34
US-09-867-550-1142
; Sequence 1142, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Mehraban, Fuad,
;  APPLICANT:  Conley, Pamela
;  APPLICANT:  Law, Debbie
;  APPLICANT:  Topper, James
;  TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
;  TITLE OF INVENTION:  Thereby
;  FILE REFERENCE: 21402-013 (Cura-313)
;  CURRENT APPLICATION NUMBER: US/09/867,550
;  CURRENT FILING DATE:  2001-09-20
;  PRIOR APPLICATION NUMBER: USSN 60/208,427
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 2125
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1142
;   LENGTH: 93
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-867-550-1142

  Query Match             3.7%;  Score 6;  DB 10;  Length 93;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     143 VTVTLP 148
           ||||||
Db      73 VTVTLP 78


RESULT 35
US-09-864-761-35148
; Sequence 35148, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35148
;   LENGTH: 99
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL022337.19
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.96
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
```

```
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
US-09-864-761-35148

  Query Match             3.7%;  Score 6;  DB 10;  Length 99;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 VGQNLV 10
           ||||||
Db       2 VGQNLV 7


RESULT 36
US-09-925-300-1428
; Sequence 1428, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
;  APPLICANT: Craig Rosen,
;  APPLICANT:  Steve Ruben
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA101
;  CURRENT APPLICATION NUMBER: US/09/925,300
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05988
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1890
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1428
;   LENGTH: 112
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (20)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (57)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1428

  Query Match             3.7%;  Score 6;  DB 10;  Length 112;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      62 TSETPR 67
           ||||||
Db       5 TSETPR 10


RESULT 37
US-09-728-914-10
; Sequence 10, Application US/09728914
; Patent No. US20010046499A1
; GENERAL INFORMATION:
;  APPLICANT: KANTOR, FRED S.
;  APPLICANT:  FIKRIG, EROL
;  APPLICANT:  DAS, SUBRATA
;  TITLE OF INVENTION: TICK ANTIGENS AND COMPOSITIONS AND METHODS COMPRISING
;  TITLE OF INVENTION:  THEM
;  FILE REFERENCE: YU-107
;  CURRENT APPLICATION NUMBER: US/09/728,914
;  CURRENT FILING DATE:  2000-12-01
;  PRIOR APPLICATION NUMBER: 60/169,048
;  PRIOR FILING DATE: 1999-12-03
;  PRIOR APPLICATION NUMBER: 60/240,716
;  PRIOR FILING DATE: 2000-10-16
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 144
;   TYPE: PRT
;   ORGANISM: Ixodes scapularis
US-09-728-914-10

  Query Match             3.7%;  Score 6;  DB 10;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     144 TVTLPD 149
           ||||||
Db     106 TVTLPD 111


RESULT 38
US-09-864-761-43168
; Sequence 43168, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43168
;   LENGTH: 182
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC004886.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
;   OTHER INFORMATION: EST_HUMAN HIT: BE967305.2, EVALUE 5.90e+00
```

```
;   OTHER INFORMATION: SWISSPROT HIT: P27110, EVALUE 3.50e-01
US-09-864-761-43168

  Query Match             3.7%;  Score 6;  DB 10;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 VVDLST 15
           ||||||
Db      47 VVDLST 52


RESULT 39
US-09-739-907-65
; Sequence 65, Application US/09739907
; Patent No. US20010012889A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 36 Human Secreted Proteins
;  FILE REFERENCE: PZ022P1
;  CURRENT APPLICATION NUMBER: US/09/739,907
;  CURRENT FILING DATE:  2000-12-20
;  PRIOR APPLICATION NUMBER: 09/348,457
;  PRIOR FILING DATE: 1999-07-07
;  PRIOR APPLICATION NUMBER: 60/070,567
;  PRIOR FILING DATE: 1998-01-07
;  PRIOR APPLICATION NUMBER: 60/070,692
;  PRIOR FILING DATE: 1998-01-07
;  PRIOR APPLICATION NUMBER: 60/070,704
;  PRIOR FILING DATE: 1998-01-07
;  PRIOR APPLICATION NUMBER: 60/070,658
;  PRIOR FILING DATE: 1998-01-07
;  NUMBER OF SEQ ID NOS: 196
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
;   LENGTH: 183
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-739-907-65

  Query Match             3.7%;  Score 6;  DB 10;  Length 183;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      82 LYLTPV 87
           ||||||
Db      17 LYLTPV 22


RESULT 40
US-10-036-869-36
; Sequence 36, Application US/10036869
; Patent No. US20020151516A1
;   GENERAL INFORMATION:
;        APPLICANT: Mixson, James A
;        TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
;                            ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
;                            THERAPY
;        NUMBER OF SEQUENCES: 43
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Connolly, Bove, Lodge, & Hutz
;             STREET: 1220 Market Street, P.O. Box 2207
;             CITY: Wilmington
;             STATE: Delaware
;             COUNTRY: U.S.A.
;             ZIP: 19899
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/036,869
;             FILING DATE: 29-No. US20020151516A1-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/985,526
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: US 08/608,845
;             FILING DATE: 16-JUL-1996
;        ATTORNEY/AGENT INFORMATION:
;             NAME: McMorrow Jr., Robert G
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (302) 658-9141
;             TELEFAX: (302) 658-5613
;   INFORMATION FOR SEQ ID NO: 36:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 185 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-10-036-869-36

  Query Match             3.7%;  Score 6;  DB 12;  Length 185;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     151 RGSVPI 156
           ||||||
Db      66 RGSVPI 71


RESULT 41
US-09-998-831-13
; Sequence 13, Application US/09998831
; Patent No. US20020119153A1
; GENERAL INFORMATION:
;  APPLICANT: Philip E. Thorpe
;  APPLICANT:  Rolf A. Brekken
;  TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
;  TITLE OF INVENTION:  INHIBITING VEGF
;  FILE REFERENCE: 4001.002584
;  CURRENT APPLICATION NUMBER: US/09/998,831
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: 09/561,108
;  PRIOR FILING DATE: 2000-04-28
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 191
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:  SYNTHETIC
US-09-998-831-13

  Query Match             3.7%;  Score 6;  DB 10;  Length 191;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     151 RGSVPI 156
           ||||||
Db      73 RGSVPI 78


RESULT 42
US-09-925-302-728
; Sequence 728, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
```

```
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 728
;    LENGTH: 192
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-09-925-302-728

  Query Match             3.7%;  Score 6;  DB 10;  Length 192;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 VLSNFS 47
           ||||||
Db     132 VLSNFS 137


RESULT 43
US-09-734-017A-74
; Sequence 74, Application US/09734017A
; Patent No. US20020142422A1
; GENERAL INFORMATION:
;  APPLICANT: Lerchl, Jens
;  APPLICANT:  Renz, Andreas
;  APPLICANT:  Ehrhardt, Thomas
;  APPLICANT:  Reindl, Andreas
;  APPLICANT:  Cirpus, Petra
;  APPLICANT:  Bischoff, Friedrich
;  APPLICANT:  Frank, Markus
;  APPLICANT:  Freund, Annette
;  APPLICANT:  Duwenig, Elke
;  APPLICANT:  Schmidt, Ralf-Michael
;  TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
;  TITLE OF INVENTION:  the
;  TITLE OF INVENTION:  synthesis of amino acids, vitamins, cofactors, nucleotides and
;  TITLE OF INVENTION:  nucleosides
;  FILE REFERENCE: BASF-NAE-1331-99-US
;  CURRENT APPLICATION NUMBER: US/09/734,017A
;  CURRENT FILING DATE:  2000-12-12
;  PRIOR APPLICATION NUMBER: 60/171,100
;  PRIOR FILING DATE: 1999-12-16
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: PatentIn Ver. 2.1/WordPerfect
; SEQ ID NO 74
;    LENGTH: 194
;    TYPE: PRT
;    ORGANISM: Physcomitrella patens
US-09-734-017A-74

  Query Match             3.7%;  Score 6;  DB 10;  Length 194;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      89 SAGGLV 94
           ||||||
Db      36 SAGGLV 41


RESULT 44
US-09-815-242-13993
; Sequence 13993, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13993
;    LENGTH: 213
;    TYPE: PRT
;    ORGANISM: Salmonella typhi
;    FEATURE:
;    NAME/KEY: VARIANT
;    LOCATION: (1)...(213)
;    OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13993

  Query Match             3.7%;  Score 6;  DB 10;  Length 213;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     145 VTLPDY 150
           ||||||
Db      26 VTLPDY 31


RESULT 45
US-09-946-239-10
; Sequence 10, Application US/09946239
; Patent No. US20020044945A1
; GENERAL INFORMATION:
;  APPLICANT: Barr, Margaret C.
;  TITLE OF INVENTION: No. US20020044945A1el Feline Immunodeficiency Virus Nucleotide
;  TITLE OF INVENTION:  Polypeptide Sequences
;  FILE REFERENCE: 18617.0059
;  CURRENT APPLICATION NUMBER: US/09/946,239
;  CURRENT FILING DATE:  2001-09-04
;  PRIOR APPLICATION NUMBER: US 09/238,303, US 60/072,927
;  PRIOR FILING DATE: 1999-01-28, 1998-01-29
;  NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 10
;    LENGTH: 252
;    TYPE: PRT
;    ORGANISM: Unknown
;    FEATURE:
;    OTHER INFORMATION: protein encoded by the vif gene of a recombinant viral
;    OTHER INFORMATION: clone constructed from the genomic DNA of a Pallas's cat felin
;    OTHER INFORMATION: immunodeficiency virus
US-09-946-239-10

  Query Match             3.7%;  Score 6;  DB 10;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      92 GLVIKA 97
           ||||||
Db     154 GLVIKA 159
```

Search completed: November 28, 2002, 19:09:50
Job time : 16 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2002, 19:09:06 ; Search time 1977 Seconds
(without alignments)
1318.904 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09900575/runat_22112002_130709_4559/app_query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09900575_@CGN_1_1_899_@runat_22112002_130709_4559 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	269.5	31.8	914	14	BQ143831	BQ143831 NF003F10D
	2	174.5	20.6	256	10	BE434026	BE434026 EST405104
	3	89.5	10.6	864	17	CNS06RF7	AL411881 T3 end of
	4	85.5	10.1	659	17	AZ568147	AZ568147 243PvF08
	5	85.5	10.1	802	12	BG666096	BG666096 DRACNG06
c	6	84	9.9	510	13	BI624785	BI624785 RH64464.5
	7	84	9.9	720	14	BQ968753	BQ968753 QHB35A23.
c	8	83.5	9.8	653	13	BJ293068	BJ293068 BJ293068
	9	83.5	9.8	657	12	BF649263	BF649263 NF055G05E
c	10	83.5	9.8	807	12	BG708796	BG708796 602673080
	11	83.5	9.8	834	12	BG621049	BG621049 602616925
	12	83.5	9.8	911	17	BH147842	BH147842 ENTPR55TR
	13	83	9.8	756	14	BQ970561	BQ970561 QHB42H01.
	14	82.5	9.7	413	12	BF557712	BF557712 UI-R-C0-h
	15	82.5	9.7	501	12	BG664251	BG664251 DRABBC11
	16	82.5	9.7	518	12	BG664658	BG664658 DRABGH01
c	17	82.5	9.7	530	12	BF072843	BF072843 219330 MA
	18	82.5	9.7	567	12	BG670398	BG670398 DRNBAG01
	19	82.5	9.7	573	12	BG671004	BG671004 DRNBKC10
c	20	82.5	9.7	602	13	BI797757	BI797757 H086C03 E
	21	82.5	9.7	696	12	BG665535	BG665535 DRACEC11
	22	82.5	9.7	697	12	BG669393	BG669393 DRNAAF07
	23	82.5	9.7	721	10	BB355470	BB355470 BB355470
	24	82.5	9.7	725	10	AW914983	AW914983 EST346287
	25	82.5	9.7	747	12	BG673676	BG673676 DRNBZB02
	26	82.5	9.7	789	12	BG671529	BG671529 DRNBTG06
c	27	82.5	9.7	1312	13	BI079789	BI079789 602876103
c	28	82	9.7	581	10	AV986132	AV986132 AV986132
c	29	82	9.7	582	10	AV996881	AV996881 AV996881
c	30	82	9.7	597	10	AV984947	AV984947 AV984947
c	31	82	9.7	614	10	AV996923	AV996923 AV996923
c	32	82	9.7	642	10	AV981815	AV981815 AV981815
c	33	82	9.7	647	10	AV974294	AV974294 AV974294
c	34	82	9.7	651	10	AV673892	AV673892 AV673892
c	35	82	9.7	689	10	AV970876	AV970876 AV970876
c	36	82	9.7	939	9	AL667637	AL667637 AL667637
	37	81.5	9.6	544	12	BG666005	BG666005 DRACMA09
c	38	81	9.6	443	9	AA969177	AA969177 op58c03.s
	39	81	9.6	594	17	BH519770	BH519770 BOGDG45TF
	40	81	9.6	669	17	AG072173	AG072173 Pan trogl
c	41	81	9.6	826	9	AL666950	AL666950 AL666950
c	42	81	9.6	868	9	AL667437	AL667437 AL667437
c	43	80.5	9.5	675	13	BI208686	BI208686 EST526726
c	44	80.5	9.5	731	13	BI208424	BI208424 EST526464
	45	80.5	9.5	779	12	BG428847	BG428847 602494632

ALIGNMENTS

RESULT 1
BQ143831/c
LOCUS BQ143831 914 bp mRNA linear EST 24-APR-2002
DEFINITION NF003F10DT1F1079 Drought Medicago truncatula cDNA clone NF003F10DT 5', mRNA sequence.
ACCESSION BQ143831
VERSION BQ143831.1 GI:20280890
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 914)

```
  AUTHORS   Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula drought library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7391
            Fax: 580 221 7380
            Email: gdmay@noble.org
            Insert Length: 914   Std Error: 0.00
            Plate: 003  row: F  column: 10
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .914
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="NF003F10DT"
                     /clone_lib="Drought"
                     /tissue_type="Plantlets"
                     /dev_stage="Pooled timepoints"
                     /note="Vector: Lambda Zap; Contains a mixture of entire
                     plantlets harvested in a series of days-post-watering
                     timepoints. "
BASE COUNT      257 a    200 c    193 g    264 t
ORIGIN

Alignment Scores:
Pred. No.:                1.09e-21      Length:          914
Score:                    269.50        Matches:         53
Percent Similarity:       65.00%        Conservative:    25
Best Local Similarity:    44.17%        Mismatches:      41
Query Match:              31.78%        Indels:          1
DB:                       14            Gaps:            1

US-09-900-575-29_COPY_26_186 (1-161) x BQ143831 (1-914)

Qy      39 TyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrPro 58
           |||         |||   ::::::   |||::::::   ::::::      :::|||||||
Db     459 TACCGGGCCTCCCTACAGTCATATAAAGGATCGCTGTACTGGAATAATGTGACCTACCCC 400

Qy      59 PheProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrp 78
           ||||||   |||:::   |||      :::            :::|||     |||
Db     399 TTCCCGTTGACGACAAATACCAACGTATTGGATATTGGTGATAAGACACCGATGCCATTA 340

Qy      79 ProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGly 98
           |||:::   ||||||:::|||||||||   :::|||||||||:::||||||||||||||||
Db     339 CCGTTAAAACTCTATATTACCCCAGTTGGGGCGGCCGGTGGAGTTGTCATAAAAGCGGGG 280

Qy      99 SerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSer---AspAspPhe 117
              :::||||||   :::   :::   :::               |||   :::
Db     279 GAAGTCATTGCCCGTATTCATATGTATAAAATTGCAACATTAGGAAGCGGAAATCCGCGT 220

Qy     118 GlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAsp 137
           |||   |||||||||   :::||||||:::||||||:::||||||||||||||||
Db     219 AATTTTACGTGGAATATTATTTCAAATAACAATGTTGTTATGCCAACAGGCGGCTGCACC 160

Qy     138 ValSerAlaArgAspValThrValThrLeuProAspTyrArgGlySerValProIlePro 157
           |||   :::|||:::|||||||||   ||||||||||:::   ||||||      ||||||
Db     159 GTTGATAGTCGAAATGTCACCGTGGATCTCCCGGACTTTCCTGGAAGCGCAGAAATTCCT 100

RESULT 2
BE434026
LOCUS       BE434026                256 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST405104 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
            clone cLEG13H23, mRNA sequence.
ACCESSION   BE434026
VERSION     BE434026.1  GI:9431869
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 256)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
            Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
            Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
            ,S.D.
  TITLE     Generation of ESTs from tomato fruit tissue, breaker stage
  JOURNAL   Unpublished (2000)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .256
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEG13H23"
                     /clone_lib="tomato breaker fruit, TIGR"
                     /tissue_type="Pericarp"
                     /dev_stage="breaker"
                     /lab_host="SOLR"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                     Site_2: XhoI; Fruit were harvested at the breaker stage
                     (first sign of lycopene accumulation on the blossom end of
                     the fruit).  Fruit were cut in half and the seeds and
                     locules were discarded prior to freezing the pericarp."
BASE COUNT       71 a     62 c     56 g     67 t
ORIGIN

Alignment Scores:
Pred. No.:                3.69e-11      Length:          256
Score:                    174.50        Matches:         35
Percent Similarity:       61.45%        Conservative:    16
Best Local Similarity:    42.17%        Mismatches:      31
Query Match:              20.58%        Indels:          1
DB:                       10            Gaps:            1

US-09-900-575-29_COPY_26_186 (1-161) x BE434026 (1-256)

Qy       6 GlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGlu 25
           ||||||||||||||||||||          |||   |||   |||||||||
Db       8 GGCCAGAATCTGGTTGTAGACTTCTGTCAGCATATAAGTTGCTGGAATGATTACGGCGGC 67

Qy      26 ThrIle---ThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSer 44
              ||||||::::::   |||   :::||||||||:::   |||   |||
Db      68 TGGTGCGACACTGATCATATAAACCTGGTACAAGGCTCAGCTTTCGCAGGCTCCCTACAG 127

Qy      45 AsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGlu 64
           ::::::   |||:::::::   ::::::      :::||||||||||||   |||:::
Db     128 TCATATAAAGGATCGCTGTACTGGAATAATGTGACCTACCCCTTCCCGTTGACGACAAAT 187

Qy      65 ThrProArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeu 84
           |||      :::            :::|||     |||   |||:::   ||||||:::
Db     188 ACCAACGTATTGGATATTGGTGATAAGACACCGATGCCATTACCGTTAAAACTCTATATT 247

Qy      85 ThrProVal 87
           ||||||||
Db     248 ACCCCAGTT 256

RESULT 3
CNS06RF7
LOCUS       CNS06RF7                864 bp    DNA     linear   GSS 05-JUL-2001
DEFINITION  T3 end of clone AW0AA013H12 of library AW0AA from strain CLIB 89 of
            Yarrowia lipolytica, genomic survey sequence.
ACCESSION   AL411881
VERSION     AL411881.1  GI:12181766
KEYWORDS    GSS.
```

```
SOURCE      Yarrowia lipolytica.
  ORGANISM  Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1  (bases 1 to 864)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 864)
  AUTHORS   Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
            lipolytica
  JOURNAL   FEBS Lett. 487 (1), 95-100 (2000)
  MEDLINE   20584727
   PUBMED   11152892
REFERENCE   3  (bases 1 to 864)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .864
                     /organism="Yarrowia lipolytica"
                     /strain="CLIB 89"
                     /db_xref="taxon:4952"
                     /clone="AW0AA013H12"
                     /clone_lib="AW0AA"
                     /note="end : T3"
     misc_feature    <1. .>854
                     /note="similar to Saccharomyces cerevisiae ORF YHR074w [
                     weak similarity to B.subtilis spore outgrowth factor B ]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
BASE COUNT      192 a    235 c    225 g    209 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:                3.77          Length:         864
Score:                    89.50         Matches:        42
Percent Similarity:       35.91%        Conservative:   23
Best Local Similarity:    23.20%        Mismatches:     51
Query Match:              10.55%        Indels:         67
DB:                       17            Gaps:           7

US-09-900-575-29_COPY_26_186 (1-161) x CNS06RF7 (1-864)

Qy       4 AsnValGlyGlnAsnLeuValValAspLeu---------SerThrGlnIlePheCys--- 19
           |||      |||:::|||      |||:::           :::::::::   |||
Db     110 AACTTCGCACAATCACTGGCTACGGATGTCTGGATCACTTTCTCGAGGGTGATCTGTATC 169

Qy      20 -------------------------HisAsnAspTyrProGluThrIleThrAspTyr 30
                                      ::::::   |||        |||   :::
Db     170 TGCACTCGTGGGAGGTCTACGCCGAGATCCTCGAACACCCCGACACCAGCGACATCATTC 229

Qy      31 ValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrVal 50
           |||||||:::      |||                  :::      |||   |||:::
Db     230 TCGACATTGGAATGCCCGTCATGC--------------ACAAGAATGTCAAGTACAATT 274

Qy      51 LysTyrSerGly-----------------------------SerSerTyrProPhePro 60
             |||                                    |||   |||   |||
Db     275 GTCGAGTCATCTCCTACAACCGGGAGATTCTGCTGATCCGACCCAAGCTGTCCCTTGCCA 334

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||      |||   |||      |||   :::            |||||||||
Db     335 ACGACGGTAACTACAGAGAGATGCGA-TACTTTACC-----------CCCTGGCCTAAG 381

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           |||   |||:::
Db     382 GCACGATACGTG---------------------------------------------- 393

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPhe--------- 117
                                                    :::|||:::
Db     394 ----------------------------------------GAAGACTACACTCTACCT 411

Qy     118 GlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAsp 137
           :::||||||   |||:::   ||||||          :::||||||   |||   |||
Db     412 CGATTTGTGCAGAATGTCTGTGCCAACGAGTCTGCTATTGTGCCGTTTGGTGATTGTGTC 471

Qy     138 ValSerAlaArgAspValThrValThrLeuProAspTyrArgGlySerValProIlePro 157
           :::|||   :::|||      :::          |||      |||:::|||      |||
Db     472 CTTTCCACCAAGGACGCCGTGATCGGTTTTGA-GACCTGTGAGGAGCTGTTCACGCCCCA 530

Qy     158 Leu 158
           :::
Db     531 GTC 533

RESULT 4
AZ568147
LOCUS       AZ568147                659 bp    DNA     linear   GSS 15-MAY-2001
DEFINITION  243PvF08 Pv MBN #30 Plasmodium vivax genomic 3', DNA sequence.
ACCESSION   AZ568147
VERSION     AZ568147.1  GI:13976954
KEYWORDS    GSS.
SOURCE      malaria parasite P. vivax.
  ORGANISM  Plasmodium vivax
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Carlton,J.M.-R. and Dame,J.B.
  TITLE     The Plasmodium vivax and P. berghei gene sequence tag projects
  JOURNAL   Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
COMMENT     Contact: Dame JB
            Dept. of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue,. Bldg 1017, Gainesville, FL 32611, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: M13(-20) forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Plasmodium vivax"
                     /strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
                     497-598)"
                     /db_xref="taxon:5855"
                     /clone_lib="Pv MBN #30"
                     /dev_stage="asexual blood forms"
                     /lab_host="Saimiri boliviensis"
                     /note="Vector: pBluescript SK(+) vector DNA, phagemid
                     excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V;
                     Host leukocytes were extracted from P. vivax infected
                     blood using the following methods: first, infected blood
                     was activated by the addition of 0.5 ml of ADP (40mg/ml)
                     per 10 ml blood. Then blood was passed over a column of
                     acid washed 0.1 mm glass beads, then through a Plasmodipur
                     filter, followed by passage through a column of pre-wet
```

Whatman CF11 powder (1:2 ratio volume of blood to CF11), and finally centrifuged through a 50% Percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 44% formamide at 50oC as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA Polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pBluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."

BASE COUNT 122 a 163 c 148 g 225 t 1 others
ORIGIN

Alignment Scores:

Pred. No.:	7.47	Length:	659
Score:	85.50	Matches:	32
Percent Similarity:	43.10%	Conservative:	18
Best Local Similarity:	27.59%	Mismatches:	32
Query Match:	10.08%	Indels:	34
DB:	17	Gaps:	6

US-09-900-575-29_COPY_26_186 (1-161) x AZ568147 (1-659)

```
Qy       9 LeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr 28
           |||       |||   |||                |||          :::      :::::::||| 
Db     130 TTGTTTAAGGACGCGAGTTTGTTCAAGTGCTGCTCCGGGAAATTCATCGCTTCTGTTACT 189

Qy      29 ------AspTyrValThrLeuGlnArgGly---------------------------- 36
                 |||   |||:::    | ||||||
Db     190 AAGGCGGATTTGGTTTCTTCATCCAGGGGGATGGTACCGAATAGGGCGTTTTCTATCCGT 249

Qy      37 ---------SerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSer 53
                  |||   :::|||||||      |||||||||   |||         :::|||
Db     250 TTCTGCGCGTCCGACTTTGGTGGGGCT---TCCAATTTTGGTGGGGTTTCCTCCTTTTCT 306

Qy      54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArg 73
           |||         |||||||||   :::         |||::::::
Db     307 GGGGACCCCCTACCCTTTCCATGTTCGCGTTCTGCCCCCAAATTG--------------- 351

Qy      74 ThrAspLysProTrpProValAlaLeuTyrLeuThr-ProVal---------------Se 88
                 |||:::|||:::      :::::::||| |||:::               ::
Db     352 ---------CCTTTTCCCATTTGGGACTTCGTCACATCCTTTATGAACAGTGTGAAGGAA 402

Qy      88 rSerAlaGlyGlyLeu------------ValIleLysAlaGlySer 99
           :|||:::||||||:::             :::|||   ||||||
Db     403 CTCCTCAGGGGGCATCAACCGAGCAGAGGAAGTTAAAATCGGTTCG 448
```

RESULT 5
BG666096
LOCUS BG666096 802 bp mRNA linear EST 30-APR-2001
DEFINITION DRACNG06 Rat DRG Library Rattus norvegicus cDNA clone DRACNG06 5', mRNA sequence.
ACCESSION BG666096
VERSION BG666096.1 GI:13888018
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 802)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
TITLE Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
MEDLINE 22056133
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
PCR PRimers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA=No.
FEATURES Location/Qualifiers
source 1. .802
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRACNG06"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/note="Total RNA was isolated from hypothalamus and transcripted into cDNA, which was then used as templat in PCR.The PCR products were separated on sequencing gel. The differential bands were cut, reamplified, cloned into pMD18-T vector and confirmed by Northern blot."
BASE COUNT 204 a 216 c 193 g 189 t
ORIGIN

Alignment Scores:

Pred. No.:	10.1	Length:	802
Score:	85.50	Matches:	29
Percent Similarity:	37.30%	Conservative:	18
Best Local Similarity:	23.02%	Mismatches:	42
Query Match:	10.08%	Indels:	37
DB:	12	Gaps:	3

US-09-900-575-29_COPY_26_186 (1-161) x BG666096 (1-802)

```
Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||         :::   |||   ||||||   |||||||:::   |||
Db     420 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC------------ 467

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db     467 ------------------------------------------------------------ 467

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg------------Val 68
                      :::|||   |||   ||||||   :::|||:::
Db     468 ---------TATCACCCACACCCTCCCACCAGCCCCAGCCCAAAAGGACAGCCCCACTCAA 518

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||       ::::::       ::::::   ||||||||||||
Db     519 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 578

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
                  ||||||   :::                ||||::::::|||        :::
Db     579 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 638

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
              :::   :::   |||   |||       |||                      :::|||
Db     639 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT-----------------CAGGATGAC 680

Qy     129 ValValValProThrGly 134
              :::||||||   |||
Db     681 ACCATTGTCCCCATAGGA 698
```

RESULT 6
BI624785/c
LOCUS BI624785 510 bp mRNA linear EST 07-SEP-2001
DEFINITION RH64464.5prime RH Drosophila melanogaster normalized Head pFlc-1 Drosophila melanogaster cDNA clone RH64464 5 similar to GlyP:

```
            FBan0007254 GO:[glycogen phosphorylase (GO:0008184); enzyme
            (GO:0003824); phosphorylase (GO:0004645)] located on: 2L 22C3-22C3
            ;: 08/23/2001, mRNA sequence.
ACCESSION   BI624785
VERSION     BI624785.1  GI:15520310
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RH Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003584: arm:2L [1824960,2149443]
            estimated-cyto:22B4-22D2: 08/23/2001
            Plate: RH.644  row: F  column: 4
            High quality sequence stop: 332.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RH64464"
                     /clone_lib="RH Drosophila melanogaster normalized Head
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="Adult"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: head; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      145 a    142 c    135 g     88 t
ORIGIN

Alignment Scores:
Pred. No.:              7.59          Length:         510
Score:                  84.00         Matches:        29
Percent Similarity:     47.62%        Conservative:   11
Best Local Similarity:  34.52%        Mismatches:     37
Query Match:            9.91%         Indels:         7
DB:                     13            Gaps:           2

US-09-900-575-29_COPY_26_186 (1-161) x BI624785 (1-510)

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg-ValValTyrAsnSe 72
           :::|||||||   |||||||:::||||||   :::      ||||| :::      |||||
Db     505 GCAGGTTCTCGATATCCATACCCAACTGGTACATGGCCTCCTCGCACTCGCTCTGAATTC 446

Qy      72 rArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGl 92
           |:::   |||           |||           |||   |||
Db     445 ACAGGTTGATCAT------GGTGTTGGTCAGTGAGCGACCCATGTAGTACTCCAGCGACA 392

Qy      92 yLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTy 112
            :::|||         |||   |||:::   |||:::::|||||||||
Db     391 GATAGTAGCCGCGCTTGGGATCCTTCTCGTAGTAGTGCTGCTGCGTCAG---------- 343

Qy     112 rAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValPr 132
            |||                |||   :::               :::||||||||||||
Db     342 -AATCCATCGGCCCACCATGTTGTCCTTGACGGTGTTGGCCAGGGTGAAGTAGTAGTCCC 284

Qy     132 oThrGlyGly 135
           |   ||||||
Db     283 TCAGGGTGGA 274

RESULT 7
BQ968753
LOCUS       BQ968753                720 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  QHB35A23.yg.ab1 QH_ABCDI sunflower RHA801 Helianthus annuus cDNA
            clone QHB35A23, mRNA sequence.
ACCESSION   BQ968753
VERSION     BQ968753.1  GI:22386274
KEYWORDS    EST.
SOURCE      common sunflower.
  ORGANISM  Helianthus annuus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
            Heliantheae; Helianthus.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
  TITLE     Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Alexander Kozik [R.W.Michelmore]
            Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QH_CA_Contig1697, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QHB35  row: A  column: 23.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Helianthus annuus"
                     /cultivar="RHA801"
                     /db_xref="taxon:4232"
                     /clone="QHB35A23"
                     /clone_lib="QH_ABCDI sunflower RHA801"
                     /lab_host="E.coli"
                     /note="Vector: pBRcDNASfiAB; The library was constructed
                     from 11 different sources of RNA from a single genotype.
                     Separate cDNAs were generated using primers that
                     incorporated unique 5' and 3' tags to distinguish each
                     source of RNA. cDNAs were then pooled, size-fractionated,
                     directionally cloned into a custom medium-copy vector and
                     transformations made with four size classes to minimize
                     size bias. Details of each source of RNA and library
                     construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_LIB=QH_ABCDI sunflower RHA801
                     TAG_TISSUE=chemical induction
                     TAG_SEQ=TGTAGCCGGG"
BASE COUNT      206 a    140 c    164 g    210 t
ORIGIN

Alignment Scores:
Pred. No.:              13            Length:         720
Score:                  84.00         Matches:        35
Percent Similarity:     43.48%        Conservative:   25
Best Local Similarity:  25.36%        Mismatches:     56
Query Match:            9.91%         Indels:         22
DB:                     14            Gaps:           6

US-09-900-575-29_COPY_26_186 (1-161) x BQ968753 (1-720)

Qy      30 TyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThr 49
           :::   ||||||||||:||   |||||||||      |||:::|||:::::|||
Db      16 TGGTACTCGATTAAACGAGCTTCTGCTTACATTATCGTGATGTCATCCTACTCGGCTTAT 75
```

```
Qy      50 ValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValVal 69
           |||:::         |||   :::         :::|||   |||:::|||
Db      76 GGAATGTACACG---CCTCAATATAAATGG---GTCAGGAACGAGCTGCCCAAAGTG--- 126

Qy      70 TyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSer 89
           :::|||||||:::  ||||||   :::   :::         |||:::   |||
Db     127 ------AACCGAACCGAGACACCATGGTTAATAGTGGTTATGCATTGTCCATTGTATAGT 180

Qy      90 AlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThr 109
           :::                 |||       :::   |||:::       :::
Db     181 AGCTATGTACATCATTATATGGAAGGTGAAACCATGAGAGTGATGTACGAACAATATTTT 240

Qy     110 AsnAsnTyrAsnSerAsp---AspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
           |||      |||      |||           :::   |||       |||
Db     241 GTTCAGTACAAAGTTGATGTTGTGTTCTCTGGACACGTTCATGCCTACGAGCGAACGGAC 300

Qy     129 ValVal----------------ValProThrGlyGlyCysAspValSerAlaArgAsp 142
           |||                 :::      |||   |||:::
Db     301 CGGGTATCGAATATAGCTTACAATATCGAAAACGGGTTGTGCAAT--------------- 345

Qy     143 ValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
                   |||   |||   ||||||   |||:::   :::|||||||
Db     346 --------------CCTAAATATGATGGATCTGCCCCTGTTTACATAACCGTT 384

RESULT 8
BJ293068/c
LOCUS       BJ293068                653 bp    mRNA    linear   EST 09-APR-2002
DEFINITION  BJ293068 Y. Ogihara unpublished cDNA library, Wh_SL Triticum
            aestivum cDNA clone whsl30h21 5', mRNA sequence.
ACCESSION   BJ293068
VERSION     BJ293068.1  GI:20108952
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 653)
  AUTHORS   Ogihara,Y. and Murai,K.
  TITLE     Expressed genes in Triticum aestivum
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .653
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="whsl30h21"
                     /clone_lib="Y. Ogihara unpublished cDNA library, Wh_SL"
                     /tissue_type="seed DPA30"
                     /dev_stage="Feekes' scale 11.3"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown under
                     hydroponic conditions at UC Davis, salt stressed  for 12
                     hours, and for 7 days, then dissected and frozen (Akhunov
                     in J  Dvorak Lab). Total RNA was prepared from sheath
                     tissue, equal quantities  of RNA were pooled from the two
                     samples, polyA was purified from the  pooled RNA, a cDNA
                     library was made, and the cDNA clones were in vivo
                     excised to give pBluescript phagemids in the TJ Close lab
                     at the  University of California, Riverside (Akhunov, Chin
                     , Choi, Close, Fenton,  Kianian, Otto, Simons, Zhang).
                     Plasmid DNA preparations and DNA  sequencing were
                     performed in the OD Anderson lab (all other authors)."
BASE COUNT      162 a    180 c    203 g    108 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                  12.8          Length:          653
Score:                      83.50         Matches:         24
Percent Similarity:         46.27%        Conservative:    7
Best Local Similarity:      35.82%        Mismatches:      17
Query Match:                9.85%         Indels:          19
DB:                         13            Gaps:            3

US-09-900-575-29_COPY_26_186 (1-161) x BJ293068 (1-653)

Qy      43 LeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThr 62
           ||||||      |||             |||         |||:::||||||:::
Db     252 CTCAGTGGCTCCGTAGGCCCCGCCAGTGGCGGTGGTCCCCGTGTACCCTGGCCCACCAGT 193

Qy      63 SerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeu 82
           :::      |||||||||                 ||||||||||||||||
Db     192 GCCAGTCGTCCCCGTGTA----------------------CCCTGGCCCGTAGCC--- 160

Qy      83 TyrLeuThrProValSerSerAlaGlyGly---------LeuValIleLysAlaGlySer 99
                           |||||||||         |||::::::   ||||||
Db     159 --------------------GCCGGCGGCCGTGGTCTGCTCCTTGTTACCGCCGGGGAG 121

Qy     100 LeuIleAlaValLeuIleLeu 106
           |||:::      |||:::|||
Db     120 CTTCTCTTTGATCTTCTGCTT 100

RESULT 9
BF649263
LOCUS       BF649263                657 bp    mRNA    linear   EST 20-DEC-2000
DEFINITION  NF055G05EC1F1038 Elicited cell culture Medicago truncatula cDNA
            clone NF055G05EC 5', mRNA sequence.
ACCESSION   BF649263
VERSION     BF649263.1  GI:11914393
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
            Center for Medicago Genomics Research
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert Length: 657   Std Error: 0.00
            Plate: 055  row: G  column: 05
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="NF055G05EC"
                     /clone_lib="Elicited cell culture"
                     /tissue_type="Cell cultures derived from root tissues"
                     /dev_stage="Cell suspensions were subcultured every 14
                     days. Cells were induced six days after subculture"
                     /note="Vector: Lambda Zap; Cells were induced with yeast
                     cell wall extracts equivalent to  50ug/ml glucose in the
                     final concentration. Samples were taken at  0.5, 1, 12 and
                     24 hours after induction. Equal amounts of RNA from  each
                     time point were pooled and used for mRNA isolation."
BASE COUNT      205 a    116 c    124 g    210 t      2 others
ORIGIN
```

```
Alignment Scores:
Pred. No.:                12.9          Length:         657
Score:                    83.50         Matches:        34
Percent Similarity:       43.22%        Conservative:   17
Best Local Similarity:    28.81%        Mismatches:     30
Query Match:              9.85%         Indels:         37
DB:                       12            Gaps:           7

US-09-900-575-29_COPY_26_186 (1-161) x BF649263 (1-657)

Qy      2 ValValAsnValGlyGlnAsnLeuValVal----------------------AspLeu 13
          ::::::|||:::|||   ||||||   :::                      ||||||
Db    349 CTGATAAATATTGGGTTGAATCTGCAAATCAATTTGATGAAGTTTGCATATCAAGATTTA 408

Qy     14 SerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeu 33
          ||||||           |||:::||||||   ||||||      |||
Db    409 TCAACA------GTGGAACATGATGATTATGTTGAAACTTATGTTGAT------------ 450

Qy     34 GlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSer 53
                    |||                 ::::::                ||||||
Db    451 --------------TACAAAGATAACTTTGATTCTTATCAAGCATTGTATGTGTATTCT 495

Qy     54 GlySerSerTyrPro------------PheProThrThrSerGluThrProArg------ 67
          |||||||||   |||            :::   ||||||:::||||||
Db    496 GGAAGTAGTGTTCCAGATTGGTTTGAGTACAAGACAACAAACGAGACAACAAACGATGAC 555

Qy     68 ---ValValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrPro 86
             :::|||      |||                          |||:::|||:::|||
Db    556 ATGATTGTTGATCTCTCT----------------------CCACTTCATCTCTCCCCT 591

Qy     87 ValSerSerAlaGlyGlyLeuValIle---------LysAlaGlySerLeuIle 101
          :::         |||   |||:::           :::   |||::::::|||
Db    592 TTGNTGGGTTTTGGTTTTTTGCTTTATCCTTCCTGAGAACGAGGGAATATGATA 645

RESULT 10
BG708796/c
LOCUS       BG708796                807 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602673080F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4796010 5',
            mRNA sequence.
ACCESSION   BG708796
VERSION     BG708796.1  GI:13986493
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 807)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10679  row: o  column: 19
            High quality sequence stop: 760.
FEATURES             Location/Qualifiers
     source          1. .807
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4796010"
                     /clone_lib="NIH_MGC_96"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.3 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length  clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      171 a    211 c    223 g    202 t
ORIGIN

Alignment Scores:
Pred. No.:                17.8          Length:         807
Score:                    83.50         Matches:        46
Percent Similarity:       42.95%        Conservative:   21
Best Local Similarity:    29.49%        Mismatches:     44
Query Match:              9.85%         Indels:         45
DB:                       12            Gaps:           8

US-09-900-575-29_COPY_26_186 (1-161) x BG708796 (1-807)

Qy     31 ValThrLeuGlnArgGly------SerAlaTyrGlyGly------ValLeuSerAsnPhe 46
          ||||||::::::   |||      |||   |||||||||      :::|||:::
Db    773 GTAACGGTGGAAAATGGAGCATACTCAGAGTACGGTGGAGGGAGAATACTTGCCCGTCAC 714

Qy     47 SerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThr------------ 62
          |||||||||          :::      :::   |||      ||||||
Db    713 TCAGGAACAGCTTAATAGACTACAAAGGCGATCCCAAGGAGTACCACGATGGTAATCAAT 654

Qy     63 ------------SerGluThr-ProArgValValTyrAsnSerArgThrAspLysProTr 78
                       |||::: ||||||      |||||||:::         |||
Db    653 CCACTCATGTTACAGGAATCCGCCGAGGACCACTTATAATAATAATCAGAGAAAGAGGCA 594

Qy     78 pProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGl 98
             |||   :::   ||||||      |||||||||   ::::::
Db    593 AAGCCGTGCTGCTTTCCAGACTCCTTCAGTTTCTGCAGGCCAAGTTCTGTATAATCTAAA 534

Qy     98 y------------------SerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTy 112
                            :::|||:::            |||::::::|||::::::
Db    533 TTATACTCCAAGCCACAAGAACCTCTTAGTACATACTGGTCTTCAGAGGACTCATAGCCT 474

Qy    112 rAsnSerAspAsp------PheGln-PheValTrpAsnIleTyrAlaAsnAsnAspValV 130
          :::|||            ||||||  ||||||   ||||||
Db    473 TCACAGCTCACCACAGTTTTTCCAAATTTGTATGCAATATC------------------ 433

Qy    130 alValProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAspT 150
                                                  |||   :::|||   |||   |
Db    432 ----------------------------------TAAGTCCGTCTTACATTCCCACTGT 408

Qy    150 yrArgGlySerValProIlePro---------LeuThrValTyr 161
          ||         :::||||||||||         ||||||:::|||
Db    407 AC------ATCATACCCATCCCAGCCTTTGTTCTGACACTGTAT 370

RESULT 11
BG621049
LOCUS       BG621049                834 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602616925F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4730406 5',
            mRNA sequence.
ACCESSION   BG621049
VERSION     BG621049.1  GI:13672420
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 834)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
```

```
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1587  row: b  column: 07
           High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .834
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4730406"
                     /clone_lib="NIH_MGC_79"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      234 a    232 c    186 g    182 t
ORIGIN

Alignment Scores:
Pred. No.:                 18.7          Length:        834
Score:                     83.50         Matches:       27
Percent Similarity:        44.33%        Conservative:  16
Best Local Similarity:     27.84%        Mismatches:    40
Query Match:               9.85%         Indels:        14
DB:                        12            Gaps:          2

US-09-900-575-29_COPY_26_186 (1-161) x BG621049 (1-834)

Qy      19 CysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAla 38
           ||||||        |||:::      |||   |||      :::::::||||||
Db     433 TGTCACCCGGGAGGACCGCAGGATCCTACACCTTACACACTCATAAAGCGAGGTGATGGG 492

Qy      39 TyrGlyGlyValLeuSerAsnPheSerGlyThrVal-LysTyrSerGlySerSerTyrPr 58
             ||||||         |||:::   |||:::   :::        |||   ||
Db     493 ACTAGAGGAGTAACTGGATATTTCACCTTCACCTTATACCTGGAGACTCCCAAGCCCTCC 552

Qy      58 oPheProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTr 78
           |   |||      :::   ||||||                    :::|||||
Db     553 ATCTCCAGCAGCAACTTAAACCCCA----------------------GGGAGGCCATG 588

Qy      78 p--------------ProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLe 93
           |              ||||||   |||   |||         :::::::::||||||
Db     589 GAAACTGTGATCTTAACCTGTGATCCTGAGACTCCGGACACAAGCTACCAGTGGTGGATG 648

Qy      93 uValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThr 109
            :::::::::|||      :::               |||   :::|||
Db     649 AATGGTCAGAGCCTCCCTTATGACTCATAGGTTTCAGCTGTCCGAAACC 697

RESULT 12
BH147842
LOCUS       BH147842                911 bp    DNA     linear   GSS 27-AUG-2001
DEFINITION  ENTPR55TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH147842
VERSION     BH147842.1  GI:15305947
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 911)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 19
            High quality sequence stop: 769.
FEATURES             Location/Qualifiers
     source          1. .911
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      325 a    115 c    182 g    289 t
ORIGIN

Alignment Scores:
Pred. No.:                 21.4          Length:        911
Score:                     83.50         Matches:       38
Percent Similarity:        43.48%        Conservative:  32
Best Local Similarity:     23.60%        Mismatches:    58
Query Match:               9.85%         Indels:        33
DB:                        17            Gaps:          7

US-09-900-575-29_COPY_26_186 (1-161) x BH147842 (1-911)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||:::   :::::::|||   :::|||   ::::::   :::||||||   :::
Db     197 CCAATCGATTCAATTGGATTTGATTTAGGACTTAAT---ACAACGCAACCATACATTATT 253

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||                        |||   :::   ||||||   :::|||
Db     254 AATGAT-------------------------ACATTTAAAATAGGATCACCATTTGGA 286

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr---ProPhe 59
           |||::::::         |||   |||      ::::::   |||   |||      |||
Db     287 GGTATGATTTATTTAAGATCTGATACAACA---TTTACAAATTCATTTTATGTCACATTC 343

Qy      60 ProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpPro 79
                          |||   :::   ||||||      |||::::::   |||
Db     344 AGTAATGTTGGAAGAGCTCCAATTATTAATTATAATATTACAACGAATGAAGAATGGAAT 403

Qy      80 ValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySer 99
                 |||                 :::|||   |||   |||   :::      :::
Db     404 AGTGTTTTA--------------AGAAATGCACCAGGAAATGTTGCAGAAATCAGAACA 448

Qy     100 LeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPhe 119
                  |||:::|||         |||      :::   :::|||   |||:::
Db     449 CCAGGAAATAGACTTGTACTTACTTCAAGAAATATTAGAAGTTTGGAAGATGCACAATAT 508

Qy     120 ValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSer 139
           :::   :::   :::                                      :::|||
Db     509 ATTAGTGATTTCTGGTTAAAA----------------------------GCAATTAGT 538

Qy     140 AlaArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThr 159
```

```
              :::      |||||||||   :::                  ||||||:::|||
Db    539 ATTTCTAATTATGCTGTTACACTTGAAAAT-----------------ATTCCAATTACA 580

Qy    160 Val 160
          :::
Db    581 TTA 583
```

RESULT 13
BQ970561
LOCUS BQ970561 756 bp mRNA linear EST 21-AUG-2002
DEFINITION QHB42H01.yg.ab1 QH_ABCDI sunflower RHA801 Helianthus annuus cDNA clone QHB42H01, mRNA sequence.
ACCESSION BQ970561
VERSION BQ970561.1 GI:22388082
KEYWORDS EST.
SOURCE common sunflower.
ORGANISM Helianthus annuus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae; Heliantheae; Helianthus.
REFERENCE 1 (bases 1 to 756)
AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegle,J., Ellison,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project http://compgenomics.ucdavis.edu/
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QH_CA_Contig1697, see http://cgpdb.ucdavis.edu/ for details.
Plate: QHB42 row: H column: 01.

```
FEATURES             Location/Qualifiers
     source          1. .756
                     /organism="Helianthus annuus"
                     /cultivar="RHA801"
                     /db_xref="taxon:4232"
                     /clone="QHB42H01"
                     /clone_lib="QH_ABCDI sunflower RHA801"
                     /lab_host="E.coli"
                     /note="Vector: pBRcDNASfiAB; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/
                     TAG_SEQ=Not found"
BASE COUNT 214 a 148 c 172 g 222 t
ORIGIN

Alignment Scores:
Pred. No.: 18.4 Length: 756
Score: 83.00 Matches: 35
Percent Similarity: 43.48% Conservative: 25
Best Local Similarity: 25.36% Mismatches: 56
Query Match: 9.79% Indels: 22
DB: 14 Gaps: 6

US-09-900-575-29_COPY_26_186 (1-161) x BQ970561 (1-756)

Qy 30 TyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThr 49
:::  :::::::::|||  |||||||||  |||:::|||:::::::|||
Db 40 TGGTACTCGATTAAACGAGCTTCTGCTTACATTATCGTGATGTCATCCTACTCGGCTTAT 99

Qy 50 ValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValVal 69
|||:::  |||  :::  :::|||  |||:::|||
Db 100 GGAATGTACACG---CCTCAATATAAATGG---CTCATGAACGAGCTGCCCAAAGTG--- 150

Qy 70 TyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSer 89
:::|||||||:::  ||||||  :::  :::  |||:::  |||
Db 151 ------AACCGAACCGAGACACCATGGTTAATAGTGGTTATGCATTGTCCATTGTATAGT 204

Qy 90 AlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThr 109
:::  |||  :::  |||:::  :::
Db 205 AGCTATGTACATCATTATATGGAAGGTGAAACCATGAGAGTGATGTACGAACAATATTTT 264

Qy 110 AsnAsnTyrAsnSerAsp---AspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
|||  |||  |||  :::  |||  |||
Db 265 GTTCAGTACAAAGTTGATGTTGTGTTCTCTGGACACGTTCATGCCTACGAGCGAACGGAC 324

Qy 129 ValVal-----ValProThrGlyGlyCysAspValSerAlaArgAsp 142
|||  :::  |||  |||:::
Db 325 CGGGTATCGAATATAGCTTACAATATCGAAAACGGGTTGTGCAAT--------------- 369

Qy 143 ValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
|||  |||  ||||||  |||:::  :::||||||
Db 370 --------------CCTAAATATGATGGATCTGCCCCTGTTTACATAACCGTT 408

RESULT 14
BF557712
LOCUS BF557712 413 bp mRNA linear EST 12-DEC-2000
DEFINITION UI-R-C0-hb-d-06-0-UI.r1 UI-R-C0 Rattus norvegicus cDNA clone UI-R-C0-hb-d-06-0-UI 5', mRNA sequence.
ACCESSION BF557712
VERSION BF557712.1 GI:11667442
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 413)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID= 1773173
Seq primer: M13 Forward.

```
FEATURES   Location/Qualifiers
     source   1. .413
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-C0-hb-d-06-0-UI"
                     /clone_lib="UI-R-C0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0 library is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged
```
```

normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C0) was constructed as follows: PCR amplified cDNA inserts from a pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the pooled UI-R-A1 and UI-R-E1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

BASE COUNT 104 a 129 c 82 g 98 t
ORIGIN

Alignment Scores:

Pred. No.:	8.28	Length:	413
Score:	82.50	Matches:	28
Percent Similarity:	37.30%	Conservative:	19
Best Local Similarity:	22.22%	Mismatches:	42
Query Match:	9.73%	Indels:	37
DB:	12	Gaps:	3

US-09-900-575-29_COPY_26_186 (1-161) x BF557712 (1-413)

```
Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||         :::   |||   ||||||   ||||||:::   |||
Db      40 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC------------ 87

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db      87 ----------------------------------------------------------- 87

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg------------Val 68
                  :::|||   |||   |||:::   :::|||:::
Db      88 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 138

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||      ::::::        ::::::   ||||||||||||
Db     139 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 198

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
                 ||||||   :::                    |||:::::::|||         :::
Db     199 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 258

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
             :::   :::   |||   |||       |||                   :::|||
Db     259 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT-----------------CAGGATGAC 300

Qy     129 ValValValProThrGly 134
              :::|||||||   |||
Db     301 ACCATTGTCCCCATAGGA 318
```

RESULT 15
BG664251
LOCUS BG664251 501 bp mRNA linear EST 30-APR-2001
DEFINITION DRABBC11 Rat DRG Library Rattus norvegicus cDNA clone DRABBC11 5', mRNA sequence.
ACCESSION BG664251
VERSION BG664251.1 GI:13886173
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 501)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L., Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
TITLE Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
MEDLINE 22056133
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
PCR PRimers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA=No.
FEATURES Location/Qualifiers
source 1. .501
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRABBC11"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/note="Total RNA was isolated from hypothalamus and transcripted into cDNA, which was then used as templat in PCR.The PCR products were separated on sequencing gel. The differential bands were cut, reamplified, cloned into pMD18-T vector and confirmed by Northern blot."
BASE COUNT 130 a 147 c 109 g 114 t 1 others
ORIGIN

Alignment Scores:

Pred. No.:	11.2	Length:	501
Score:	82.50	Matches:	28
Percent Similarity:	37.30%	Conservative:	19
Best Local Similarity:	22.22%	Mismatches:	42
Query Match:	9.73%	Indels:	37
DB:	12	Gaps:	3

US-09-900-575-29_COPY_26_186 (1-161) x BG664251 (1-501)

```
Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||         :::   |||   ||||||   ||||||:::   |||
Db     189 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC------------ 236

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db     236 ----------------------------------------------------------- 236

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg------------Val 68
                  :::|||   |||   |||:::   :::|||:::
Db     237 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 287

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||      ::::::        ::::::   ||||||||||||
Db     288 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 347

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
                 ||||||   :::                    |||:::::::|||         :::
Db     348 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 407

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
             :::   :::   |||   |||       |||                   :::|||
Db     408 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT-----------------CAGGATGAC 449
```

```
Qy    129 ValValValProThrGly 134
          :::||||||   |||
Db    450 ACCATTGTCCCCATAGGA 467

RESULT 16
BG664658
LOCUS       BG664658                518 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRABGH01 Rat DRG Library Rattus norvegicus cDNA clone DRABGH01 5',
            mRNA sequence.
ACCESSION   BG664658
VERSION     BG664658.1  GI:13886580
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.
            , Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
            ,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRABGH01"
                     /clone_lib="Rat DRG Library"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /note="Total RNA was isolated from hypothalamus and
                     transcripted into cDNA,  which was then used as templat in
                     PCR.The PCR products were separated  on sequencing gel.
                     The differential bands were cut, reamplified, cloned into
                     pMD18-T vector and confirmed by Northern blot."
BASE COUNT      132 a    150 c    112 g    124 t
ORIGIN

Alignment Scores:
Pred. No.:                11.8          Length:          518
Score:                    82.50         Matches:         28
Percent Similarity:       37.30%        Conservative:    19
Best Local Similarity:    22.22%        Mismatches:      42
Query Match:              9.73%         Indels:          37
DB:                       12            Gaps:            3

US-09-900-575-29_COPY_26_186 (1-161) x BG664658 (1-518)

Qy     13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
          |||      :::   |||   ||||||   ||||||:::   |||
Db    189 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC---------- 236

Qy     33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db    236 ------------------------------------------------------------ 236

Qy     53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg----------Val 68
                  :::|||   |||   |||:::   :::|||:::
Db    237 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 287

Qy     69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
          |||   :::|||      ::::::         ::::::   ||||||||||||
Db    288 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 347

Qy     89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
          ||||||   :::                 |||:::::::|||          :::
Db    348 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 407

Qy    109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
          :::   :::   |||   |||   |||                    :::|||
Db    408 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT----------CAGGATGAC 449

Qy    129 ValValValProThrGly 134
          :::||||||   |||
Db    450 ACCATTGTCCCCATAGGA 467

RESULT 17
BF072843/c
LOCUS       BF072843                530 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  219330 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF072843
VERSION     BF072843.1  GI:10866354
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 77  row: B  column: 2
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT      126 a    131 c    182 g     91 t
```

```
ORIGIN

Alignment Scores:
Pred. No.:                12.2          Length:          530
Score:                    82.50         Matches:         26
Percent Similarity:       55.38%        Conservative:    10
Best Local Similarity:    40.00%        Mismatches:      20
Query Match:              9.73%         Indels:          9
DB:                       12            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x BF072843 (1-530)

Qy      47 SerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrPro 66
           ||||||   |||   |||   |||:::|||:::      ||||||   :::      |||
Db     419 TCAGGTCTGGTTTTGTATGTGGGCACCTCACACTGTGCCCCGACCCCCAAC------CCT 366

Qy      67 ArgVal---------ValTyrAsnSerArgThrAspLysPro---------TrpProVal 80
           |||         ||||||  |||||||||   ||||||          ||||||
Db     365 GCAGTCACCATGCCTGTGTACCCCTCCAGAACACAGAAGCCTCAGATGCAATGGCCTCCC 306

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           |||   :::   :::      |||::::::   |||   |||          |||   |||
Db     305 GCCTGGCACAGGTCTCAAGCCTCAGCCTCAGCAGGT---GTTAGGGGTGGTGGGGAGTTG 249

Qy     101 IleAlaValLeuIle 105
           ::::::   |||:::
Db     248 CTGAGCCCTCTCCTT 234

RESULT 18
BG670398
LOCUS       BG670398                 567 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNBAG01 Rat DRG Library Rattus norvegicus cDNA clone DRNBAG01 5',
            mRNA sequence.
ACCESSION   BG670398
VERSION     BG670398.1  GI:13892497
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.
            , Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
            ,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNBAG01"
                     /clone_lib="Rat DRG Library"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /note="Total RNA was isolated from hypothalamus and
                     transcripted into cDNA,  which was then used as templat in
                     PCR.The PCR products were separated  on sequencing gel.
                     The differential bands were cut, reamplified, cloned into
                     pMD18-T vector and confirmed by Northern blot."
BASE COUNT      134 a    165 c    130 g    138 t
ORIGIN

Alignment Scores:
Pred. No.:                13.5          Length:          567
Score:                    82.50         Matches:         28
Percent Similarity:       37.30%        Conservative:    19
Best Local Similarity:    22.22%        Mismatches:      42
Query Match:              9.73%         Indels:          37
DB:                       12            Gaps:            3

US-09-900-575-29_COPY_26_186 (1-161) x BG670398 (1-567)

Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||         :::   |||   ||||||   ||||||:::   |||
Db      67 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC------------ 114

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db     114 ---------------------------------------------------------- 114

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg------------Val 68
                   :::|||   |||   |||:::   :::|||:::
Db     115 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 165

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||      ::::::         ::::::   ||||||||||||
Db     166 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 225

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
                ||||||   :::                 |||:::::::|||         :::
Db     226 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 285

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
              :::   :::   |||   |||      |||                  :::|||
Db     286 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT-----------------CAGGATGAC 327

Qy     129 ValValValProThrGly 134
           :::||||||   |||
Db     328 ACCATTGTCCCCATAGGA 345

RESULT 19
BG671004
LOCUS       BG671004                 573 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNBKC10 Rat DRG Library Rattus norvegicus cDNA clone DRNBKC10 5',
            mRNA sequence.
ACCESSION   BG671004
VERSION     BG671004.1  GI:13893103
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 573)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.
            , Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
            ,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
```

```
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .573
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNBKC10"
                     /clone_lib="Rat DRG Library"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /note="Total RNA was isolated from hypothalamus and
                     transcripted into cDNA,  which was then used as templat in
                     PCR.The PCR products were separated  on sequencing gel.
                     The differential bands were cut, reamplified, cloned into
                     pMD18-T vector and confirmed by Northern blot."
BASE COUNT      145 a    161 c    124 g    142 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                13.8          Length:          573
Score:                    82.50         Matches:         28
Percent Similarity:       37.30%        Conservative:    19
Best Local Similarity:    22.22%        Mismatches:      42
Query Match:              9.73%         Indels:          37
DB:                       12            Gaps:            3

US-09-900-575-29_COPY_26_186 (1-161) x BG671004 (1-573)

Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||         :::   |||   ||||||   |||||||:::   |||
Db     134 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC------------ 181

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db     181 ------------------------------------------------------------ 181

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg------------Val 68
                  :::|||   |||   |||:::   :::|||:::
Db     182 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 232

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||      ::::::         ::::::   ||||||||||||
Db     233 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 292

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
                 ||||||   :::                  |||:::::|||         :::
Db     293 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 352

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
              :::   :::   |||   |||      |||                    :::|||
Db     353 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT-----------------CAGGATGAC 394

Qy     129 ValValValProThrGly 134
              :::|||||   |||
Db     395 ACCATTGTCCCCATAGGA 412

RESULT 20
BI797757/c
LOCUS       BI797757                602 bp    mRNA    linear   EST 02-OCT-2001
DEFINITION  H086C03 Endosperm library from Oryza sativa (10 days after anthesis
            ) Oryza sativa cDNA clone H086C03, mRNA sequence.
ACCESSION   BI797757
VERSION     BI797757.1  GI:15849481
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X., Wu
            ,H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.
  TITLE     A Gene Expression Screen in Oryza sativa
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dong HT
            Laboratory of Functional Genetics
            Bio-technology Institute of Zhejiang University
            Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
            Tel: 0086-571-86892051
            Fax: 0086-571-86961525
            Email: htdong@zjuem.zju.edu.cn
            Seq primer: M13 forward primer.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /clone="H086C03"
                     /clone_lib="Endosperm library from Oryza sativa (10 days
                     after anthesis)"
                     /tissue_type="Endosperm"
                     /dev_stage="10 days after anthesis"
                     /note="Vector: pSport2"
BASE COUNT      170 a    145 c    147 g    140 t
ORIGIN

Alignment Scores:
Pred. No.:                14.9          Length:          602
Score:                    82.50         Matches:         22
Percent Similarity:       44.74%        Conservative:    12
Best Local Similarity:    28.95%        Mismatches:      31
Query Match:              9.73%         Indels:          11
DB:                       13            Gaps:            3

US-09-900-575-29_COPY_26_186 (1-161) x BI797757 (1-602)

Qy      60 ProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpPro 79
           |||||||||   :::   |||   |||:::                    :::||||||||||
Db     215 CCTACCACGGGTGATCCACCACCTGTGATG-----------------CGCCCTTGGCCA 174

Qy      80 ValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySer 99
                 |||:::|||   |||                :::      |||   |||
Db     173 GCGGTCACGTACATCACCGGTGTGTGCGTTTACCGTCCAATGAGGAGTAAGCAAGGGCAT 114

Qy     100 LeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPhe 119
              :::                    :::      :::   ||||||   ::::::
Db     113 TATTTGGAG-----------AACGATTCTGGTGGCACTCATCTGGATGATATTGAGTAT 66

Qy     120 ValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGly 135
              |||   |||   |||          |||:::::|||   ||||||
Db      65 ---TGGGAAATTCTGGCTATTGGCCTTGTGATCCTTCCACATGGGGGT 21

RESULT 21
BG665535
LOCUS       BG665535                696 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRACEC11 Rat DRG Library Rattus norvegicus cDNA clone DRACEC11 5',
            mRNA sequence.
ACCESSION   BG665535
VERSION     BG665535.1  GI:13887457
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
```

```
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.
            , Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
            ,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRACEC11"
                     /clone_lib="Rat DRG Library"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /note="Total RNA was isolated from hypothalamus and
                     transcripted into cDNA,  which was then used as templat in
                     PCR.The PCR products were separated  on sequencing gel.
                     The differential bands were cut, reamplified, cloned into
                     pMD18-T vector and confirmed by Northern blot."
BASE COUNT      168 a    187 c    162 g    174 t      5 others
ORIGIN

Alignment Scores:
Pred. No.:              18.6          Length:        696
Score:                  82.50         Matches:       28
Percent Similarity:     37.30%        Conservative:  19
Best Local Similarity:  22.22%        Mismatches:    42
Query Match:            9.73%         Indels:        37
DB:                     12            Gaps:          3

US-09-900-575-29_COPY_26_186 (1-161) x BG665535 (1-696)

Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||         :::   |||   ||||||   |||||||:::   |||
Db     120 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC------------ 167

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db     167 ------------------------------------------------------------ 167

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg------------Val 68
                   :::|||   |||   |||:::   :::|||:::
Db     168 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 218

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||      ::::::         ::::::   ||||||||||||
Db     219 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 278

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
                 ||||||   :::                  |||::::::|||         :::
Db     279 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 338

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128

                 :::   :::   |||   |||      |||                    :::|||
Db     339 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT-----------------CAGGATGAC 380

Qy     129 ValValValProThrGly 134
           :::||||||   |||
Db     381 ACCATTGTCCCCATAGGA 398

RESULT 22
BG669393
LOCUS       BG669393                 697 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNAAF07 Rat DRG Library Rattus norvegicus cDNA clone DRNAAF07 5',
            mRNA sequence.
ACCESSION   BG669393
VERSION     BG669393.1  GI:13891315
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.
            , Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
            ,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .697
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNAAF07"
                     /clone_lib="Rat DRG Library"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /note="Total RNA was isolated from hypothalamus and
                     transcripted into cDNA,  which was then used as templat in
                     PCR.The PCR products were separated  on sequencing gel.
                     The differential bands were cut, reamplified, cloned into
                     pMD18-T vector and confirmed by Northern blot."
BASE COUNT      173 a    189 c    159 g    169 t      7 others
ORIGIN

Alignment Scores:
Pred. No.:              18.7          Length:        697
Score:                  82.50         Matches:       28
Percent Similarity:     37.30%        Conservative:  19
Best Local Similarity:  22.22%        Mismatches:    42
Query Match:            9.73%         Indels:        37
DB:                     12            Gaps:          3

US-09-900-575-29_COPY_26_186 (1-161) x BG669393 (1-697)

Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
```

```
          |||          :::   |||   ||||||   ||||||:::   |||
Db     84 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC------------ 131

Qy     33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db    131 ------------------------------------------------------------ 131

Qy     53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg------------Val 68
                  :::|||   |||   |||:::   :::|||:::
Db    132 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 182

Qy     69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
          |||   :::|||     ::::::     ::::::   ||||||||||||
Db    183 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 242

Qy     89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
             ||||||   :::              |||:::::|||         :::
Db    243 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 302

Qy    109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
             :::   :::   |||   |||      |||                  :::|||
Db    303 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT----------------CAGGATGAC 344

Qy    129 ValValValProThrGly 134
             :::|||||||   |||
Db    345 ACCATTGTCCCCATAGGA 362
```

RESULT 23
BB355470

```
LOCUS       BB355470                721 bp    mRNA    linear   EST 24-OCT-2001
DEFINITION  BB355470 RIKEN full-length enriched, adult male corpus striatum Mus
            musculus cDNA clone C030006F24 3', mRNA sequence.
ACCESSION   BB355470
VERSION     BB355470.2  GI:16405712
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 12, 2000 this sequence version replaced gi:9067298.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES                Location/Qualifiers
     source             1. .721
                        /organism="Mus musculus"
                        /strain="C57BL/6J"
                        /db_xref="taxon:10090"
                        /clone="C030006F24"
                        /clone_lib="RIKEN full-length enriched, adult male corpus
                        striatum"
                        /sex="male"
                        /tissue_type="corpus striatum"
                        /dev_stage="adult"
                        /lab_host="DH10B"
                        /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN. Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full-length by
                        cap-trapper. cDNA went through one round of normalization
                        to Rot = 10.0 and subtraction to Rot = 185.0. Second
                        strand cDNA was prepared with the primer adapter of
                        sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                        3']. cDNA was cloned into the XhoI and BamHI sites.
                        Vector: a modified pBluescript KS(+) after bulk excision
                        from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                        BamHI"
BASE COUNT      190 a    213 c    146 g    172 t
ORIGIN
```

Alignment Scores:

Pred. No.:	19.7	Length:	721
Score:	82.50	Matches:	32
Percent Similarity:	41.38%	Conservative:	16
Best Local Similarity:	27.59%	Mismatches:	51
Query Match:	9.73%	Indels:	17
DB:	10	Gaps:	5

US-09-900-575-29_COPY_26_186 (1-161) x BB355470 (1-721)

```
Qy      4 AsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyr 23
          |||   |||   |||:::          |||          |||          ||||||
Db    320 AATGCCGGCTCCAACATCACGCGCCTGCTGCTGGCTTCCCGCTTT---------GACTAC 370

Qy     24 Pro------GluThrIleThrAspTyrVal----------------------ThrLeu 33
                ::::|||:::   ||||||                          |||
Db    371 AGCAGCCTCGATACCGTCTGGGACTACCAGCTGCTTGTCCACATTACTGATGACAACTTG 430

Qy     34 GlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSer 53
             ||||||              |||         :::|||||||||      |||
Db    431 CTGTCTGGCAGCACAAAAGCCAAGGCTCTTGTTGAA---ACAGGGACAGTGACTCTGAGT 487

Qy     54 GlySerSerTyrProPheProThrThrSerGluThr---------ProArgValValTyr 70
                  |||   |||||||||:::   |||         ||||||:::   |||
Db    488 GTTAAAGTCATTCCTCACCCAACCACAACCATCACAACTCCCAGGCCCAGGATTACCTAC 547
```

```
Qy      71 AsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAla 90
           |||   :::      :::      :::   :::      |||
Db     548 CAGATCCGCAGGGAAAACGTGTACTCCACATCTGCATGGTATGTGCCGTTCATCGTCACT 607

Qy      91 GlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeu 106
           |||   ::::::      |||   |||   :::::::::::|||
Db     608 CTGGGATCCATACTTCTTCTTGGCCTTCTGGGGTCCCTGATGGTGCTG 655

RESULT 24
AW914983
LOCUS       AW914983                725 bp    mRNA    linear   EST 25-MAY-2000
DEFINITION  EST346287 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone
            RGIBN16 5' end, mRNA sequence.
ACCESSION   AW914983
VERSION     AW914983.1  GI:8080667
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            This clone is available through the ATCC, contact the ATCC
            tel#703-365-2700 for further information
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="RGIBN16"
                     /clone_lib="Normalized rat ovary, Bento Soares"
                     /note="Organ: ovary; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
BASE COUNT      182 a    200 c    166 g    177 t
ORIGIN

Alignment Scores:
Pred. No.:                19.8          Length:          725
Score:                    82.50         Matches:         28
Percent Similarity:       37.30%        Conservative:    19
Best Local Similarity:    22.22%        Mismatches:      42
Query Match:              9.73%         Indels:          37
DB:                       10            Gaps:            3

US-09-900-575-29_COPY_26_186 (1-161) x AW914983 (1-725)

Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||         :::   |||   ||||||   ||||||:::   |||
Db     311 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC---------- 358

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db     358 ------------------------------------------------------------ 358

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg----------Val 68
                  :::|||   |||   |||:::   :::|||:::
Db     359 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 409

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||      ::::::         ::::::   ||||||||||||
Db     410 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 469

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
           ||||||   :::                |||:::::|||          :::
Db     470 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 529

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
             :::   :::   |||   |||      |||                  :::|||
Db     530 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT----------CAGGATGAC 571

Qy     129 ValValValProThrGly 134
           :::||||||   |||
Db     572 ACCATTGTCCCCATAGGA 589

RESULT 25
BG673676
LOCUS       BG673676                747 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNBZB02 Rat DRG Library Rattus norvegicus cDNA clone DRNBZB02 5',
            mRNA sequence.
ACCESSION   BG673676
VERSION     BG673676.1  GI:13895775
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.
            , Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
            ,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNBZB02"
                     /clone_lib="Rat DRG Library"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /note="Total RNA was isolated from hypothalamus and
                     transcripted into cDNA,  which was then used as templat in
                     PCR.The PCR products were separated  on sequencing gel.
                     The differential bands were cut, reamplified, cloned into
                     pMD18-T vector and confirmed by Northern blot."
BASE COUNT      197 a    201 c    164 g    181 t      4 others
ORIGIN

Alignment Scores:
Pred. No.:                20.8          Length:          747
Score:                    82.50         Matches:         28
Percent Similarity:       37.30%        Conservative:    19
Best Local Similarity:    22.22%        Mismatches:      42
```

```
Query Match:          9.73%      Indels:        37
DB:                   12         Gaps:          3

US-09-900-575-29_COPY_26_186 (1-161) x BG673676 (1-747)

Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||          :::   |||   ||||||   |||||||:::   |||
Db     109 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC------------ 156

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db     156 ------------------------------------------------------------ 156

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg------------Val 68
                     :::|||   |||   |||:::   :::|||:::
Db     157 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 207

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||     ::::::          ::::::   ||||||||||||
Db     208 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 267

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
                 ||||||   :::          |||::::::|||         :::
Db     268 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 327

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
              :::   :::   |||   |||      |||                  :::|||
Db     328 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT----------------CAGGATGAC 369

Qy     129 ValValValProThrGly 134
              :::||||||   |||
Db     370 ACCATTGTCCCCATAGGA 387
```

RESULT 26
BG671529

```
LOCUS       BG671529                789 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNBTG06 Rat DRG Library Rattus norvegicus cDNA clone DRNBTG06 5',
            mRNA sequence.
ACCESSION   BG671529
VERSION     BG671529.1  GI:13893628
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 789)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.
            , Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
            ,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .789
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNBTG06"
                     /clone_lib="Rat DRG Library"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /note="Total RNA was isolated from hypothalamus and
                     transcripted into cDNA,  which was then used as templat in
                     PCR.The PCR products were separated  on sequencing gel.
                     The differential bands were cut, reamplified, cloned into
                     pMD18-T vector and confirmed by Northern blot."
BASE COUNT      193 a    213 c    192 g    191 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:              22.6          Length:          789
Score:                  82.50         Matches:         28
Percent Similarity:     37.30%        Conservative:    19
Best Local Similarity:  22.22%        Mismatches:      42
Query Match:            9.73%         Indels:          37
DB:                     12            Gaps:            3

US-09-900-575-29_COPY_26_186 (1-161) x BG671529 (1-789)

Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||          :::   |||   ||||||   |||||||:::   |||
Db     164 TTACAACTGCTTGTCCTCTGTGCAAACGATGCCCCGGAAAGTTAGACC------------ 211

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52

Db     211 ------------------------------------------------------------ 211

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg------------Val 68
                     :::|||   |||   |||:::   :::|||:::
Db     212 ---------TATCACCCACACCCTCCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 262

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||     ::::::          ::::::   ||||||||||||
Db     263 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 322

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
                 ||||||   :::          |||::::::|||         :::
Db     323 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 382

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
              :::   :::   |||   |||      |||                  :::|||
Db     383 GAATCAAAATTCTGCTCAGTAGACGTCTCCTTT----------------CAGGATGAC 424

Qy     129 ValValValProThrGly 134
              :::||||||   |||
Db     425 ACCATTGTCCCCATAGGA 442
```

RESULT 27
BI079789/c

```
LOCUS       BI079789                1312 bp    mRNA    linear   EST 20-JUN-2001
DEFINITION  602876103F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5007745 5',
            mRNA sequence.
ACCESSION   BI079789
VERSION     BI079789.1  GI:14498119
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1312)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
```

```
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11051  row: f  column: 02
           High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1. .1312
                     /organism="Mus musculus"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5007745"
                     /clone_lib="NCI_CGAP_Mam2"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      274 a    431 c    410 g    197 t
ORIGIN

Alignment Scores:
Pred. No.:                49.8          Length:          1312
Score:                    82.50         Matches:         35
Percent Similarity:       39.47%        Conservative:    10
Best Local Similarity:    30.70%        Mismatches:      52
Query Match:              9.73%         Indels:          17
DB:                       13            Gaps:            5

US-09-900-575-29_COPY_26_186 (1-161) x BI079789 (1-1312)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPhe------SerGlyThrValLysTyrSerGlySer 55
           |||      ||||||      |||         ||||||||         |||
Db     809 GCCGTTCGTGGCGTCACAAAGGGCTTTAAGCTTGGGGGAACCGTCCTCCGTGTGGGC--- 753

Qy      56 SerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArgThrAsp 75
               |||      :::        :::   |||      ::::::||||||
Db     752 ------CCCCTTTCTTCCAACCAACGCTCTGTACGCCCAAAGTTCTCCTCGCGA---CCA 702

Qy      76 LysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIle 95
           |||||||||         |||  :::||||||||||   |||   |||   ::::::
Db     701 AAACCTTGGGCTTCCTACCTGGGCGTTACCCCTGTGCCTTCGCGTGGCGCGGTTATACCA 642

Qy      96 LysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAsp 115
           ||||||                              :::   :::||||||
Db     641 AAGGCT------------------------------AGGGCGTCCAATTACCGCGTCGTG 612

Qy     116 AspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGly 135
                   |||                  |||  :::     |||   |||
Db     611 CAATGTGTCCCTGGTTGGGGCTCGACCGGAGGCAACCGGCTGGAATTTCCTTGGGGCTGT 552

Qy     136 CysAspValSerAlaArgAspValThrValThr---LeuPro 148
           |||         |||      |||      |||   ||||||
Db     551 TGTGGCCAGTGTGCCGTCCCGGTTCTGGGAACCCATCTGCCT 510

RESULT 28
AV986132/c
LOCUS       AV986132               581 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  AV986132 Nori Satoh unpublished cDNA library, larva Ciona
            intestinalis cDNA clone cilv39p09 5', mRNA sequence.
ACCESSION   AV986132
VERSION     AV986132.1  GI:19475000
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="cilv39p09"
                     /clone_lib="Nori Satoh unpublished cDNA library, larva"
                     /tissue_type="whole animal"
                     /dev_stage="larva"
                     /note="Vector: pBluescript SK"
BASE COUNT      146 a    179 c    114 g    142 t
ORIGIN

Alignment Scores:
Pred. No.:                16.1          Length:          581
Score:                    82.00         Matches:         36
Percent Similarity:       40.31%        Conservative:    16
Best Local Similarity:    27.91%        Mismatches:      61
Query Match:              9.67%         Indels:          17
DB:                       10            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AV986132 (1-581)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||   ||||||      ||||||          :::      |||
Db     409 GCTGATGGCGGGCGGGCATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 350

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer---------------- 72
              :::              :::|||      :::|||
Db     349 GACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTG 290

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||      |||   ||||||          ||||||   :::      ||||||
Db     289 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 230

Qy      93 LeuVal---------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                        |||   |||:::               |||
Db     229 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 171

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
                |||         :::   :::|||     ||||||   :::   |||
Db     170 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 111

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::          |||   |||:::::|||         |||   |||:::
Db     110 GACAAGATT---------GGTTGGTGTGAATTGTCCGGTTCCTTGGCAGATGACTTGTGG 60

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           :::::::|||   |||   |||   |||
Db      59 ATTTCCTTGGATGACGTAAGGGCCTCG 33

RESULT 29
AV996881/c
LOCUS       AV996881               582 bp    mRNA    linear   EST 15-MAR-2002
DEFINITION  AV996881 Nori Satoh unpublished cDNA library, larva Ciona
            intestinalis cDNA clone cilv29n02 5', mRNA sequence.
ACCESSION   AV996881
VERSION     AV996881.1  GI:19488215
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
```

```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="cilv29n02"
                     /clone_lib="Nori Satoh unpublished cDNA library, larva"
                     /tissue_type="whole animal"
                     /dev_stage="larva"
                     /note="Vector: pBluescript SK"
BASE COUNT      146 a    178 c    114 g    144 t
ORIGIN

Alignment Scores:
Pred. No.:                 16.2          Length:          582
Score:                     82.00         Matches:         36
Percent Similarity:        40.31%        Conservative:    16
Best Local Similarity:     27.91%        Mismatches:      61
Query Match:               9.67%         Indels:          17
DB:                        10            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AV996881 (1-582)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||   ||||||      ||||||          :::      |||
Db     411 GCTGATGGCGGGCGGGCATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 352

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer-------------- 72
              :::             :::|||      :::|||
Db     351 GACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTG 292

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||       |||   ||||||         ||||||   :::      ||||||
Db     291 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 232

Qy      93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                        |||   |||:::                |||
Db     231 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 173

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
               |||          :::  :::|||      ||||||   :::   |||
Db     172 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 113

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::        |||   |||::::::|||         |||   |||:::
Db     112 GACAAGATT---------GGTTGGTGTGAATTGTCCGGTTCCTTGGCAGATGACTTGTGG 62

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           ::::::|||   |||   |||   |||
Db      61 ATTTCCTTGGATGACGTAAGGGCCTCG 35

RESULT 30
AV984947/c
LOCUS       AV984947                597 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  AV984947 Nori Satoh unpublished cDNA library, larva Ciona
            intestinalis cDNA clone cilv40d02 5', mRNA sequence.
ACCESSION   AV984947
VERSION     AV984947.1  GI:19473815
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis

  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="cilv40d02"
                     /clone_lib="Nori Satoh unpublished cDNA library, larva"
                     /tissue_type="whole animal"
                     /dev_stage="larva"
                     /note="Vector: pBluescript SK"
BASE COUNT      149 a    184 c    117 g    147 t
ORIGIN

Alignment Scores:
Pred. No.:                 16.8          Length:          597
Score:                     82.00         Matches:         36
Percent Similarity:        40.31%        Conservative:    16
Best Local Similarity:     27.91%        Mismatches:      61
Query Match:               9.67%         Indels:          17
DB:                        10            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AV984947 (1-597)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||   ||||||      ||||||          :::      |||
Db     415 GCTGATGGCGGGCGGGCATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 356

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer-------------- 72
              :::             :::|||      :::|||
Db     355 GACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTA 296

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||       |||   ||||||         ||||||   :::      ||||||
Db     295 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 236

Qy      93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                        |||   |||:::                |||
Db     235 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 177

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
               |||          :::  :::|||      ||||||   :::   |||
Db     176 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 117

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::        |||   |||::::::|||         |||   |||:::
Db     116 GACAAGATT---------GGTTGGTGTGAATTGTCCGGTTCCTTGGCAGATGACTTGTGG 66

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           ::::::|||   |||   |||   |||
Db      65 ATTTCCTTGGATGACGTAAGGGCCTCG 39

RESULT 31
AV996923/c
LOCUS       AV996923                614 bp    mRNA    linear   EST 15-MAR-2002
DEFINITION  AV996923 Nori Satoh unpublished cDNA library, larva Ciona
            intestinalis cDNA clone cilv29p03 5', mRNA sequence.
ACCESSION   AV996923
VERSION     AV996923.1  GI:19488257
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
```

```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .614
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="cilv29p03"
                     /clone_lib="Nori Satoh unpublished cDNA library, larva"
                     /tissue_type="whole animal"
                     /dev_stage="larva"
                     /note="Vector: pBluescript SK"
BASE COUNT      158 a    185 c    121 g    150 t
ORIGIN

Alignment Scores:
Pred. No.:                 17.6          Length:          614
Score:                     82.00         Matches:         36
Percent Similarity:        40.31%        Conservative:    16
Best Local Similarity:     27.91%        Mismatches:      61
Query Match:               9.67%         Indels:          17
DB:                        10            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AV996923 (1-614)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||   ||||||      ||||||          :::     |||
Db     443 GCTGATGGCGGGCGGGCATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 384

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer--------------- 72
              :::             :::|||     :::|||
Db     383 GACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTG 324

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||      |||   ||||||        ||||||   :::      ||||||
Db     323 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 264

Qy      93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                        |||   |||:::                |||
Db     263 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 205

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
               |||       :::   :::|||      ||||||   :::   |||
Db     204 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 145

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::         |||   |||::::::|||         |||   |||:::
Db     144 GACAAGATT---------GGTTGGTGTGAATTGTCCGGTTCCTTGGCAGATGACTTGTGG 94

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           ::::::|||   |||   |||   |||
Db      93 ATTTCCTTGGATGACGTAAGGGCCTCG 67

RESULT 32
AV981815/c
LOCUS       AV981815                642 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  AV981815 Nori Satoh unpublished cDNA library, larva Ciona
            intestinalis cDNA clone cilv36c06 5', mRNA sequence.
ACCESSION   AV981815
VERSION     AV981815.1  GI:19471214
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
```

```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="cilv36c06"
                     /clone_lib="Nori Satoh unpublished cDNA library, larva"
                     /tissue_type="whole animal"
                     /dev_stage="larva"
                     /note="Vector: pBluescript SK"
BASE COUNT      160 a    200 c    125 g    157 t
ORIGIN

Alignment Scores:
Pred. No.:                 18.8          Length:          642
Score:                     82.00         Matches:         36
Percent Similarity:        40.31%        Conservative:    16
Best Local Similarity:     27.91%        Mismatches:      61
Query Match:               9.67%         Indels:          17
DB:                        10            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AV981815 (1-642)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||   ||||||      ||||||          :::     |||
Db     393 GCTGATGGCGGGCGGGCATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 334

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer--------------- 72
              :::             :::|||     :::|||
Db     333 GACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTA 274

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||      |||   ||||||        ||||||   :::      ||||||
Db     273 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 214

Qy      93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                        |||   |||:::                |||
Db     213 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 155

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
               |||       :::   :::|||      ||||||   :::   |||
Db     154 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 95

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::         |||   |||::::::|||         |||   |||:::
Db      94 GACAAGATT---------GGTTGGTGTGAATTGTCCGGTTCCTTGGCAGATGACTTGTGG 44

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           ::::::|||   |||   |||   |||
Db      43 ATTTCCTTGGATGACGTAAGGGCCTCG 17

RESULT 33
AV974294/c
LOCUS       AV974294                647 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  AV974294 Nori Satoh unpublished cDNA library, larva Ciona
            intestinalis cDNA clone cilv34f08 5', mRNA sequence.
ACCESSION   AV974294
VERSION     AV974294.1  GI:19464058
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
```

```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="cilv34f08"
                     /clone_lib="Nori Satoh unpublished cDNA library, larva"
                     /tissue_type="whole animal"
                     /dev_stage="larva"
                     /note="Vector: pBluescript SK"
BASE COUNT      170 a    203 c    121 g    153 t
ORIGIN

Alignment Scores:
Pred. No.:                19.1          Length:          647
Score:                    82.00         Matches:         36
Percent Similarity:       40.31%        Conservative:    16
Best Local Similarity:    27.91%        Mismatches:      61
Query Match:              9.67%         Indels:          17
DB:                       10            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AV974294 (1-647)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||   ||||||      ||||||          :::      |||
Db     629 GCTGATGGCGGGCGGGCATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 570

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer-------------- 72
             :::             :::|||      :::|||
Db     569 GACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTA 510

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||      |||   ||||||         ||||||  :::      ||||||
Db     509 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 450

Qy      93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                        |||   |||:::              |||
Db     449 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 391

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
               |||      :::   :::|||      ||||||  :::   |||
Db     390 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 331

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::         |||   |||::::::|||         |||   |||:::
Db     330 GACAAGATT---------GGTTGGTGTGAATTGTCCGGTTCCTTGGCAGATGACTTGTGG 280

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           ::::::|||   |||   |||   |||
Db     279 ATTTCCTTGGATGACGTAAGGGCCTCG 253

RESULT 34
AV673892/c
LOCUS       AV673892                651 bp    mRNA    linear   EST 05-OCT-2000
DEFINITION  AV673892 Nori Satoh unpublished cDNA library Ciona intestinalis
            cDNA clone citb11e8 5', mRNA sequence.
ACCESSION   AV673892
VERSION     AV673892.1  GI:10111891
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
```

```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="citb11e8"
                     /clone_lib="Nori Satoh unpublished cDNA library"
                     /tissue_type="whole animal"
                     /dev_stage="tailbud"
                     /note="Vector: pBluescript SK"
BASE COUNT      176 a    207 c    122 g    146 t
ORIGIN

Alignment Scores:
Pred. No.:                19.3          Length:          651
Score:                    82.00         Matches:         36
Percent Similarity:       40.31%        Conservative:    16
Best Local Similarity:    27.91%        Mismatches:      61
Query Match:              9.67%         Indels:          17
DB:                       10            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AV673892 (1-651)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||   ||||||      ||||||          :::      |||
Db     562 GCTGATGGCGGGCGGGCATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 503

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer-------------- 72
             :::             :::|||      :::|||
Db     502 GACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTA 443

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||      |||   ||||||         ||||||  :::      ||||||
Db     442 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 383

Qy      93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                        |||   |||:::              |||
Db     382 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 324

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
               |||      :::   :::|||      ||||||  :::   |||
Db     323 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 264

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::         |||   |||::::::|||         |||   |||:::
Db     263 GACAAGATT---------GGTTGGTGTGAATTGTCCGGTTCCTTGGCAGATGACTTGTGG 213

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           ::::::|||   |||   |||   |||
Db     212 ATTTCCTTGGATGACGTAAGGGCCTCG 186

RESULT 35
AV970876/c
LOCUS       AV970876                689 bp    mRNA    linear   EST 14-MAR-2002
DEFINITION  AV970876 Nori Satoh unpublished cDNA library, larva Ciona
            intestinalis cDNA clone cilv19n17 5', mRNA sequence.
ACCESSION   AV970876
VERSION     AV970876.1  GI:19460640
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.
  TITLE     Expressed genes in Ciona intestinalis
```

```
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Nori Satoh
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4081
            Fax: 81-75-705-1113
            Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .689
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="cilv19n17"
                     /clone_lib="Nori Satoh unpublished cDNA library, larva"
                     /tissue_type="whole animal"
                     /dev_stage="larva"
                     /note="Vector: pBluescript SK"
BASE COUNT      185 a    215 c    132 g    154 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:                21              Length:         689
Score:                    82.00           Matches:        36
Percent Similarity:       40.31%          Conservative:   16
Best Local Similarity:    27.91%          Mismatches:     61
Query Match:              9.67%           Indels:         17
DB:                       10              Gaps:           4

US-09-900-575-29_COPY_26_186 (1-161) x AV970876 (1-689)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||   ||||||      ||||||           :::      |||
Db     580 GCTGATGGCGGGCGGGCATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 521

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer---------------- 72
             :::             :::|||     :::|||
Db     520 GACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTA 461

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||      |||   ||||||         ||||||   :::      ||||||
Db     460 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 401

Qy      93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                        |||   |||:::                |||
Db     400 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 342

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
               |||          :::   :::|||      ||||||   :::   |||
Db     341 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 282

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::         |||   |||::::::|||         |||   |||:::
Db     281 GACAAGATT---------GGTTGGTGTGAATTGTCCGGTTCCTTGGCAGATGACTTGTGG 231

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           ::::::|||   |||   |||   |||
Db     230 ATTTCCTTGGATGACGTAAGGGCCTCG 204

RESULT 36
AL667637/c
LOCUS       AL667637               939 bp    mRNA    linear   EST 11-JAN-2002
DEFINITION  AL667637 directional larval cDNA library Ciona intestinalis cDNA
            clone 022ZC11 5', mRNA sequence.
ACCESSION   AL667637
VERSION     AL667637.1  GI:18134544
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 939)
  AUTHORS   Genoscope.
  TITLE     Ciona intestinalis directional larval cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            IMPORTANT: this sequence may contain errors. The Ciona intestinalis
            library from which the clone was isolated may be contaminated with
            cDNAs from bacteria or other Eukarya.
            Directional larval cDNA library originate from Dr.M.Branno,
            Stazione A.Dohrn, Naples, Italy, and was prepared in
            pBluescript2SK+.
FEATURES             Location/Qualifiers
     source          1. .939
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="022ZC11"
                     /clone_lib="directional larval cDNA library"
                     /note="Vector: pBluescript2SK+"
BASE COUNT      247 a    279 c    187 g    225 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                34              Length:         939
Score:                    82.00           Matches:        36
Percent Similarity:       40.31%          Conservative:   16
Best Local Similarity:    27.91%          Mismatches:     61
Query Match:              9.67%           Indels:         17
DB:                       9               Gaps:           4

US-09-900-575-29_COPY_26_186 (1-161) x AL667637 (1-939)

Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||   ||||||      ||||||           :::      |||
Db     771 GCTGATGGCGGTCGGACATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 712

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer---------------- 72
             :::             :::|||     :::|||
Db     711 GACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTA 652

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||      |||   ||||||         ||||||   :::      ||||||
Db     651 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 592

Qy      93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                        |||   |||:::                |||
Db     591 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 533

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
               |||          :::   :::|||      ||||||   :::   |||
Db     532 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 473

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::         |||   |||::::::|||         |||   |||:::
Db     472 GACAAGATT---------GGTTGGTGTGAATTGTCCAGTTCCTTGGCAGATGACTTGTGG 422

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           ::::::|||   |||   |||   |||
Db     421 ATTTCCTTGGATGACGTAAGGGCCTCG 395

RESULT 37
BG666005
LOCUS       BG666005               544 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRACMA09 Rat DRG Library Rattus norvegicus cDNA clone DRACMA09 5',
            mRNA sequence.
ACCESSION   BG666005
VERSION     BG666005.1  GI:13887927
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 544)
```

```
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.
            , Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
            ,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRACMA09"
                     /clone_lib="Rat DRG Library"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /note="Total RNA was isolated from hypothalamus and
                     transcripted into cDNA,  which was then used as templat in
                     PCR.The PCR products were separated  on sequencing gel.
                     The differential bands were cut, reamplified, cloned into
                     pMD18-T vector and confirmed by Northern blot."
BASE COUNT      144 a    147 c    124 g    129 t
ORIGIN

Alignment Scores:
Pred. No.:                16.7          Length:          544
Score:                    81.50         Matches:         28
Percent Similarity:       38.10%        Conservative:    20
Best Local Similarity:    22.22%        Mismatches:      41
Query Match:              9.61%         Indels:          37
DB:                       12            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x BG666005 (1-544)

Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||         :::   |||   ||||||   |||
Db     225 TTACAACTGCTTGTGCTCTGTTCAAACGATGCCCCG---------------------- 260

Qy      33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52
                                                 |||:::      |||
Db     261 ------------------------------------------GGAAGTTAGACCTAT 275

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArg-----------Val 68
                  :::|||   |||   |||:::   :::|||:::
Db     276 -----------CACCCACACCCTGCCACCACCCCCAGCCCAAAAGGACAGCCCACTCAA 323

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::|||      ::::::         ::::::   ||||||||||||
Db     324 GTCTCTTCTTCCACAGTGAACCGTGGTTTGTTACTAAGTACGTATTTGACACCAACAGTT 383

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
                |||||   :::                |||::::::|||        :::
Db     384 CTAACTGGTGGAACGATTAGATCTCGCACACTAAGTATTAGCATCCTAACTCACGACCGA 443

Qy     109 ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAsp 128
              :::   :::   |||   |||                 :::|||      :::|||

Db     444 GAATCAAAATTCTGCTCAGTAGAC----------------GTCTACTTTCAGGATGAC 485

Qy     129 ValValValProThrGly 134
              :::||||||   |||
Db     486 ACCATTGTCCCCATAGGA 503

RESULT 38
AA969177/c
LOCUS       AA969177                 443 bp    mRNA    linear   EST 07-JUL-1998
DEFINITION  op58c03.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1581028 3' similar to gb:U02570 !!!! ALU CLASS C WARNING
            ENTRY !!!! (HUMAN);, mRNA sequence.
ACCESSION   AA969177
VERSION     AA969177.1  GI:3144357
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Insert Length: 603   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1581028"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      122 a     83 c    136 g    102 t
ORIGIN

Alignment Scores:
Pred. No.:                14            Length:          443
Score:                    81.00         Matches:         28
Percent Similarity:       43.75%        Conservative:    14
Best Local Similarity:    29.17%        Mismatches:      44
Query Match:              9.55%         Indels:          10
DB:                       9             Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AA969177 (1-443)

Qy      64 GluThrProArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyr 83
           ::::::|||:::   :::      :::   |||   ||||||          |||
Db     328 GACTCTCCCCAGACCATC------AACAGTACATCCAAACCCCTGAGCCCTGGGCTGCTG 275

Qy      84 LeuThrProValSerSerAlaGly--------------GlyLeuValIleLysAlaGly 98
           |||          |||   |||                |||||||:::
Db     274 CTGATTCTTCATCAGTCCCTTGGTCTCTAGCCTCGGCCAGGGCTGCTCTGCTCTCCGCCC 215

Qy      99 SerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGln 118
```

```
          |||   ::::::   |||   |||      |||:::   :::|||   |||   |||
Db    214 TCCTTTGTGTCTCAACTGACGCTC------ACCGACCCATTCAATGGAGACCCTTTCGCA 161

Qy    119 PheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspVal 138
            :::|||:::                    |||      |||      ||||||
Db    160 TCTATTTGGTCTTTCTCATGTCCCCCACTGGTCTGTACCCCAGGGAGCGGGTGCTTGTAC 101

Qy    139 SerAlaArgAspVal---ThrValThrLeuProAspTyrArgGlySer 153
                      |||   |||   ||||||   |||:::   |||:::
Db    100 TGTGTGAATCCAGTGTTCACATTCACACTTAATGACTTCCTTGGCACC 53

RESULT 39
BH519770
LOCUS       BH519770                 594 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGDG45TF BOGD Brassica oleracea genomic clone BOGDG45, DNA
            sequence.
ACCESSION   BH519770
VERSION     BH519770.1  GI:17727855
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGDG45TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGDG45"
                     /clone_lib="BOGD"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      125 a    174 c    144 g    151 t
ORIGIN

Alignment Scores:
Pred. No.:                 22          Length:          594
Score:                     81.00       Matches:         25
Percent Similarity:        46.07%      Conservative:    16
Best Local Similarity:     28.09%      Mismatches:      34
Query Match:               9.55%       Indels:          14
DB:                        17          Gaps:            2

US-09-900-575-29_COPY_26_186 (1-161) x BH519770 (1-594)

Qy     24 ProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeu 43
          |||   |||   |||                  |||   :::||||||::::::
Db    305 CCGTCCACGGCTACCTCTTCCCGTCGTTTCAACGATTCTTCGACGTTCGGCGGTTTGATC 364

Qy     44 SerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhe---------Pro 60
          |||       :::   :::|||   ::::::   |||      |||:::         |||
Db    365 TCGAAATTGAACCGATCGGTGCCGTTCACGTCATCGATTTGACCGTACGGTGTCATACCT 424

Qy     61 ThrThrSerGluThrProArgValValTyrAsn-SerArgThr-AspLysProTrpProV 80
             |||:::       :::       |||    :::|||      ||||||
Db    425 AATACGAATCCGATGAGGAAGAAAGCGAGGAATTGGTAAAACGGTCGTCGCCATGG---- 480

Qy     80 alAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerL 100
                                                  :::|||||||||||      |||:::   |||::::
Db    481 ----------------------ACCTCTGCGGGAGGCTTTGGGATCAGAGGAGGAGCTG 517

Qy    100 euIleAlaValLeuIleLeuArg 107
           ::   |||      :::||||||
Db    518 TGTTTGCCGCTGCTCTTCTGAGA 540

RESULT 40
AG072173
LOCUS       AG072173                 669 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-063J11.F, genomic survey sequence.
ACCESSION   AG072173
VERSION     AG072173.1  GI:16623975
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-063J11.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 669)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-063J11.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      155 a    166 c    135 g    213 t
ORIGIN

Alignment Scores:
Pred. No.:                 26.5        Length:          669
Score:                     81.00       Matches:         42
Percent Similarity:        37.93%      Conservative:    24
Best Local Similarity:     24.14%      Mismatches:      55
Query Match:               9.55%       Indels:          53
DB:                        17          Gaps:            10

US-09-900-575-29_COPY_26_186 (1-161) x AG072173 (1-669)

Qy      2 ValValAsnValGlyGlnAsnLeu----------ValValAspLeuSer 14
          |||   |||   ||||||::::::          :::|||   ::::::
Db    145 GTTAGAAACCAAGGCCAGGATGTGACACTTGAATCTGGACTCTACCTTGTATCTGTCAAT 204

Qy     15 ThrGlnIlePheCys----------HisAsnAspTyrProGluThrIleThr 28
          :::|||:::                |||:::   :::|||
Db    205 TCTCAACTGGCCGCTATCTCATCTTTAGCCTAACACTCACCTTGGCCTGCTAATACAAAC 264

Qy     29 AspTyrValThrLeuGlnArg---GlySerAlaTyrGlyGlyValLeuSerAsnPheSer 47
             :::||||||||||   |||   |||||||||              ::::::|||
```

```
Db    265 CCATGGGTTACACTAATGAGGCTGGGCTCTGCTACATCCTGTACCCAGGCTTCATTCCAT 324

Qy     48 GlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArg 67
          |||   |||   |||   |||   :::|||   |||                    :::
Db    325 GGGATAGTGCCTTATATTGGAGATAATTATTCTTTTGTGGGAGCCAAGAATATTCAGCAA 384

Qy     68 ValValTyrAsnSerArgThrAspLysProTrpPro-----------ValAlaLeu--- 82
          |||:::                  ||||||           |||   |||
Db    385 AGTGTTTGG-------------------TGGCCCTCAGTAAACAGAGTTCAGCTAGTT 423

Qy     83 ---------TyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySer 99
               .:::|||   |||                  |||:::           |||
Db    424 CTTTGGTCCTTCCTGTGTCCA--------------AAACTTATTTTCCTCTGTGCTTCC 468

Qy    100 LeuIleAlaVal-LeuIleLeuArgGlnThrAsnAsnTyrAsnSerAsp---------- 115
             :::   ||| |||::::::         ||||||
Db    469 TGCCTTCCTGTCTTTGCTAATACTAACTTTAAACAACGTGGGCCAGAAAATTCAGCTTCT 528

Qy    116 -----------------------------------AspPheGlnPheValTrpAsnIl 123
                                               |||||||:::   ||||||   ::
Db    529 TAAATCTCCCAGGGTCCATGATTATCTGGCTCTCCGAGATTTTGAA---GTCTGGCCTGT 585

Qy    123 eTyrAlaAsnAsnAspValValValProThrGlyGlyCys 136
          :   :::|||   |||   :::              |||
Db    586 A---AGCAACAAGGACTTTCTTACCATGGCTGCATGCTGT 622

RESULT 41
AL666950/c
LOCUS       AL666950                826 bp    mRNA    linear   EST 11-JAN-2002
DEFINITION  AL666950 directional larval cDNA library Ciona intestinalis cDNA
            clone 013ZC09 5', mRNA sequence.
ACCESSION   AL666950
VERSION     AL666950.1  GI:18133857
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 826)
  AUTHORS   Genoscope.
  TITLE     Ciona intestinalis directional larval cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            IMPORTANT: this sequence may contain errors. The Ciona intestinalis
            library from which the clone was isolated may be contaminated with
            cDNAs from bacteria or other Eukarya.
            Directional larval cDNA library originate from Dr.M.Branno,
            Stazione A.Dohrn, Naples, Italy, and was prepared in
            pBluescript2SK+.
FEATURES             Location/Qualifiers
     source          1. .826
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="013ZC09"
                     /clone_lib="directional larval cDNA library"
                     /note="Vector: pBluescript2SK+"
BASE COUNT      204 a    247 c    160 g    213 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:                36.7          Length:          826
Score:                    81.00         Matches:         36
Percent Similarity:       40.31%        Conservative:    16
Best Local Similarity:    27.91%        Mismatches:      61
Query Match:              9.55%         Indels:          17
DB:                       9             Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AL666950 (1-826)

Qy     38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
          |||   ||||||       ||||||            :::       |||
Db    440 GCTGATGGCGGGCGGACATCTAATACGCATGATGTAATTTTGGTTTCAAATTGACCGTTG 381

Qy     58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer-------------- 72
             :::                 :::|||      :::|||
Db    380 AACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTA 321

Qy     73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
          :::   |||       |||   ||||||         ||||||   :::       ||||||
Db    320 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 261

Qy     93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
          ||||||                        |||   |||:::           |||
Db    260 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 202

Qy    108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
                |||           :::   :::|||      ||||||   :::   |||
Db    201 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 142

Qy    128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
          |||   :::         |||   |||::::::|||         |||   |||:::
Db    141 GACAAGATT---------GGTTGGTGTGAATTGTCCGGTTCCTTGGCAGATGACTTGTGG 91

Qy    145 ValThrLeuProAspTyrArgGlySer 153
          ::::::|||   |||   |||   |||
Db     90 ATTTCCTTGGATGACGTAAGGGCCTCG 64

RESULT 42
AL667437/c
LOCUS       AL667437                868 bp    mRNA    linear   EST 11-JAN-2002
DEFINITION  AL667437 directional larval cDNA library Ciona intestinalis cDNA
            clone 019ZE10 5', mRNA sequence.
ACCESSION   AL667437
VERSION     AL667437.1  GI:18134344
KEYWORDS    EST.
SOURCE      Ciona intestinalis.
  ORGANISM  Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 868)
  AUTHORS   Genoscope.
  TITLE     Ciona intestinalis directional larval cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            IMPORTANT: this sequence may contain errors. The Ciona intestinalis
            library from which the clone was isolated may be contaminated with
            cDNAs from bacteria or other Eukarya.
            Directional larval cDNA library originate from Dr.M.Branno,
            Stazione A.Dohrn, Naples, Italy, and was prepared in
            pBluescript2SK+.
FEATURES             Location/Qualifiers
     source          1. .868
                     /organism="Ciona intestinalis"
                     /db_xref="taxon:7719"
                     /clone="019ZE10"
                     /clone_lib="directional larval cDNA library"
                     /note="Vector: pBluescript2SK+"
BASE COUNT      224 a    269 c    168 g    204 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:                39.7          Length:          868
Score:                    81.00         Matches:         36
Percent Similarity:       40.31%        Conservative:    16
Best Local Similarity:    27.91%        Mismatches:      61
Query Match:              9.55%         Indels:          17
DB:                       9             Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x AL667437 (1-868)
```

```
Qy      38 AlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyr 57
           |||  ||||||      ||||||            :::      |||
Db     586 GCTGATGGCGGTCGGACATCTAATACGCATGATGTAATTTTGGTNTCAAATTGACCGTTG 527

Qy      58 ProPheProThrThrSerGluThrProArgValValTyrAsnSer-------------- 72
             :::              :::|||      :::|||
Db     526 CACTGGCAGATTGTGAAAGTGTCTCCGGCAACGATATATCCAGGTTGGCATTGGTATGTA 467

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           :::   |||       |||   ||||||          ||||||   :::      ||||||
Db     466 AAAGCTGATCCAACCTGGTACGTGGCGGACGAAGAAACGCCAAATGCTGGTACAGGGGGG 407

Qy      93 LeuVal--------------IleLysAlaGlySerLeuIleAlaValLeuIleLeuArg 107
           ||||||                    |||   |||:::              |||
Db     406 TTGGTGCAAGTAACTACATTAACTACGGGTGGTGGGTTAGTAGTCACTCGAACAAG-CGT 348

Qy     108 GlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsn 127
               |||         :::   :::|||      ||||||   :::   |||
Db     347 GCGGAAGGGAACTGTTGGGACAACTGGTGGCAATGTGACTGGAACTCTTGGCAGAATACA 288

Qy     128 AspValValValProThrGlyGlyCysAspValSer---------AlaArgAspValThr 144
           |||   :::         |||   |||:::::|||         |||   |||:::
Db     287 GACAAGATT---------GGTTGGTGTGAATTGTCCAGTTCCTTGGCAGATGACTTGTGG 237

Qy     145 ValThrLeuProAspTyrArgGlySer 153
           ::::::|||   |||   |||   |||
Db     236 ATTTCCTTGGATGACGTAAGGGCCTCG 210

RESULT 43
BI208686/c
LOCUS       BI208686                675 bp    mRNA    linear   EST 11-JUL-2001
DEFINITION  EST526726 cTOS Lycopersicon esculentum cDNA clone cTOS17H8 5' end,
            mRNA sequence.
ACCESSION   BI208686
VERSION     BI208686.1  GI:14686410
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,
            Ronning,C. and Tanksley,S.
  TITLE     Generation of ESTs from Tomato Suspension Cultures
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .675
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496, E6203"
                     /db_xref="taxon:4081"
                     /clone="cTOS17H8"
                     /clone_lib="cTOS"
                     /tissue_type="suspension cultures"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Suspension cultures of L.esculentum E6203 were grown
                     in Murashige and Skoog based medium, supplemented with 15%
                     coconut milk (filter sterilized and added after
                     autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
                     Fresh medium was added every 7 days, and cultures were
                     grown at 25 C, with 12hrs of light and continuous
                     shaking."
BASE COUNT      222 a     97 c    177 g    179 t
ORIGIN

Alignment Scores:

Pred. No.:                30.8          Length:          675
Score:                    80.50         Matches:         33
Percent Similarity:       35.62%        Conservative:    19
Best Local Similarity:    22.60%        Mismatches:      53
Query Match:              9.49%         Indels:          41
DB:                       13            Gaps:            5

US-09-900-575-29_COPY_26_186 (1-161) x BI208686 (1-675)

Qy      15 ThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGln 34
           |||          ||||||   :::::::|||:::|||   |||   |||      :::|||
Db     633 ACTAACTACATTTGCCATCCAAATCATCCAAAAACTACTACTTCTTACTTTTTTGTTCAA 574

Qy      35 ArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGly 54
           :::   |||       |||   ::::::                                  
Db     573 CAAACGTCTATAAGGGGGGAAATCATT---------------------------------- 547

Qy      55 SerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArgThr 74
                        ||||||                  :::                 :::
Db     546 --------------CCAACTCAAATTTTCCGAACTCTTCTGGCTGTAGTTCTCCTCTCT 502

Qy      75 AspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuVal 94
           |||   |||      :::::::::   :::                      ||||||   :::
Db     501 GATAACCCCACAGATTTGTCCATGCTGATT-----------GGTAAAGGAGGCTTCCTG 454

Qy      95 IleLysAlaGlySerLeuIleAlaValLeu-----------------------IleLeu 106
                     ||||||:::|||||||||                        :::|||
Db     453 TTC---------TCCTTACTAGCTGTGCTTCTCAAGAAATACCTTGAAGTTTCTCTCCTC 403

Qy     107 ArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsn 126
                                 :::   ||||||      :::|||||||||
Db     402 TCTGCTATTGGTGTAGAAGCAACAACTGATTTCCTAGCGCTTTGGAACATCTTA------ 349

Qy     127 AsnAspValValValProThrGlyGlyCysAspValSerAlaArgAspValThrValThr 146
                              |||                  |||   |||:::|||
Db     348 ----------------------GGCTCTCCTTTTTTCTTGGAAGATTCAACTCTAACT 313

Qy     147 LeuProAspTyrArgGly 152
                |||   |||
Db     312 CCCAAATCATACACTGGT 295

RESULT 44
BI208424/c
LOCUS       BI208424                731 bp    mRNA    linear   EST 11-JUL-2001
DEFINITION  EST526464 cTOS Lycopersicon esculentum cDNA clone cTOS17E9 5' end,
            mRNA sequence.
ACCESSION   BI208424
VERSION     BI208424.1  GI:14686148
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 731)
  AUTHORS   van der Hoeven,R., Sun,H., Bezzerides,J., Cho,J., Utterback,R.,
            Ronning,C. and Tanksley,S.
  TITLE     Generation of ESTs from Tomato Suspension Cultures
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .731
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496, E6203"
                     /db_xref="taxon:4081"
                     /clone="cTOS17E9"
                     /clone_lib="cTOS"
                     /tissue_type="suspension cultures"
```

```
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Suspension cultures of L.esculentum E6203 were grown
                     in Murashige and Skoog based medium, supplemented with 15%
                     coconut milk (filter sterilized and added after
                     autoclaving), 2% saccharose, and 1mg/ml 2,4D (pH5.8).
                     Fresh medium was added every 7 days, and cultures were
                     grown at 25 C, with 12hrs of light and continuous
                     shaking."
BASE COUNT      235 a    108 c    192 g    196 t
ORIGIN

Alignment Scores:
Pred. No.:                 34.9          Length:         731
Score:                     80.50         Matches:        33
Percent Similarity:        35.62%        Conservative:   19
Best Local Similarity:     22.60%        Mismatches:     53
Query Match:               9.49%         Indels:         41
DB:                        13            Gaps:           5

US-09-900-575-29_COPY_26_186 (1-161) x BI208424 (1-731)

Qy      15 ThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGln 34
           |||          ||||||   ::::::|||:::|||   |||   |||      :::|||
Db     633 ACTAACTACATTTGCCATCCAAATCATCCAAAAACTACTACTTCTTACTTCTTTGTTCAA 574

Qy      35 ArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGly 54
           :::   |||      |||   ::::::
Db     573 CAAACGTCTATAAGGGGGGAAATCATT------------------------------ 547

Qy      55 SerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArgThr 74
                    ||||||                 :::                     :::
Db     546 --------------CCAACTCAAATTTTCCGAACTCTTCTGGCTGTAGTTCTCCTCTCT 502

Qy      75 AspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuVal 94
           |||   |||      :::::::::  :::              ||||||   :::
Db     501 GATAACCCCACAGATTTGTCCATGCTGATT-----------GGTAAAGGAGGCTTCCTG 454

Qy      95 IleLysAlaGlySerLeuIleAlaValLeu----------------------IleLeu 106
                   |||||||:::|||||||||                        :::|||
Db     453 TTC---------TCCTTACTAGCTGTGCTTCTCAAGAAATACCTTGAAGTTTCTCTCCTC 403

Qy     107 ArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsn 126
                        :::   ||||||      :::||||||||||
Db     402 TCTGCTATTGGTGTAGAAGCAACAACTGATTTCCTAGCGCTTTGGAACATCTTA------ 349

Qy     127 AsnAspValValValProThrGlyGlyCysAspValSerAlaArgAspValThrValThr 146
                              |||                   |||   |||:::|||
Db     348 ----------------------GGCTCTCCTTTTTTCTTGGAAGATTCAACTCTAACT 313

Qy     147 LeuProAspTyrArgGly 152
                 |||   |||
Db     312 CCCAAATCATACACTGGT 295

RESULT 45
BG428847
LOCUS       BG428847                 779 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602494632F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4608631 5',
            mRNA sequence.
ACCESSION   BG428847
VERSION     BG428847.1  GI:13335353
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 779)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1350  row: p  column: 08
            High quality sequence stop: 673.
FEATURES             Location/Qualifiers
     source          1. .779
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4608631"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      212 a    166 c    204 g    197 t
ORIGIN

Alignment Scores:
Pred. No.:                 38.5          Length:         779
Score:                     80.50         Matches:        31
Percent Similarity:        36.84%        Conservative:   18
Best Local Similarity:     23.31%        Mismatches:     31
Query Match:               9.49%         Indels:         53
DB:                        12            Gaps:           6

US-09-900-575-29_COPY_26_186 (1-161) x BG428847 (1-779)

Qy      13 LeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
           |||:::      :::   |||   :::      |||   :::         |||:::|||
Db     131 CTGACTCCCTATTTAACTTGCCCCAGCTCGGCTCCTTCTTCCCAGGCCTGCTACATTACC 190

Qy      33 Leu------GlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrVal 50
           |||      :::::::|||   |||      |||:::   ||||||
Db     191 CTCCATGATGAAAAGGGAGGGGCTGAGACAGGTCTGTGTTCAAAT-------------- 235

Qy      51 LysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyr 70
                              |||      |||||||||   :::      |||
Db     236 --------------------CCAAGCTCCACCACTTCTTCCAGTGAATCAGTGAGCATA 274

Qy      71 AsnSerArg----------------------------ThrAspLysProTrpProVal 80
           |||                                  |||:::|||:::|||
Db     275 AATGGAGAGCACTCTAAGGGCCCCAAAACACTGTGAGTGGAAGATGAACCCTACCCTCAA 334

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100

Db     334 ---------------------------------------------------------- 334

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPhe--- 119
              |||||||||:::         :::      :::||||||||||:::   ||||||
Db     335 ---GCAGTGCTGGTC---------TCCCAGTGATTTAACTCTGATGAATCCCAGTTTACC 382

Qy     120 ------------ValTrpAsnIleTyrAlaAsnAsnAsp 128
                       :::|||||||:::         :::|||
Db     383 CGTTCTGGTAAAATCTGGAATCTGAAGGGTATGGATGAT 421


Search completed: November 28, 2002, 20:31:09
Job time : 1986 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2002, 19:06:11 ; Search time 257 Seconds
(without alignments)
1410.785 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09900575/runat_22112002_130708_4535/app_query.fasta_1.327
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=45
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09900575_@CGN_1_1_125_@runat_22112002_130708_4535 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	848	100.0	837	24	AAD29358	Escherichia coli s
2	841	99.2	837	24	AAD29368	Escherichia coli s
3	841	99.2	900	16	AAQ93062	FimH protein gene
4	834	98.3	837	24	AAD29357	Escherichia coli s
5	834	98.3	837	24	AAD29372	Escherichia coli s
6	834	98.3	840	24	AAD29361	Escherichia coli s
7	834	98.3	840	24	AAD29363	Escherichia coli s
8	834	98.3	900	16	AAQ93070	FimH protein gene
9	834	98.3	900	16	AAQ93071	FimH protein gene
10	834	98.3	903	21	AAA66015	E. coli proliferat
11	834	98.3	903	24	ABK11187	DNA encoding Esche
12	834	98.3	903	24	AAD29376	Escherichia coli s
13	834	98.3	4637	24	ABK23921	Vector pFIMAFGH ex
14	834	98.3	7416	24	AAD29374	Plasmid pCGA126-1
15	834	98.3	9299	24	ABK23922	Vector pFIMAICFGH
16	831	98.0	837	24	AAD29382	Escherichia coli s
17	830	97.9	837	24	AAD29354	Escherichia coli s
18	830	97.9	900	16	AAQ93074	FimH protein gene
19	826	97.4	837	24	AAD29366	Escherichia coli s
20	825	97.3	840	24	AAD29356	Escherichia coli s
21	825	97.3	900	16	AAQ93067	FimH protein gene
22	825	97.3	900	16	AAQ93073	FimH protein gene
23	825	97.3	900	16	AAQ93068	FimH protein gene
24	824	97.2	837	24	AAD29371	Escherichia coli s
25	824	97.2	837	24	AAD29510	Escherichia coli s
26	822	96.9	837	24	AAD29367	Escherichia coli s
27	822	96.9	837	24	AAD29369	Escherichia coli s
28	822	96.9	900	16	AAQ93063	FimH protein gene
29	820	96.7	837	24	AAD29359	Escherichia coli s
30	819	96.6	837	24	AAD29362	Escherichia coli s
31	819	96.6	837	24	AAD29373	Escherichia coli s
32	819	96.6	837	24	AAD29509	Escherichia coli s
33	819	96.6	840	24	AAD29355	Escherichia coli s
34	818	96.5	837	24	AAD29360	Escherichia coli s
35	817	96.3	837	24	AAD29352	Escherichia coli s
36	816	96.2	837	24	AAD29364	Escherichia coli s
37	816	96.2	837	24	AAD29365	Escherichia coli s
38	816	96.2	1227	24	ABL52756	Fusion protein prt
39	815	96.1	900	16	AAQ93066	FimH protein gene
40	815	96.1	900	16	AAQ93072	FimH protein gene
41	814	96.0	900	16	AAQ93065	FimH protein gene
42	814	96.0	900	16	AAQ93075	FimH protein gene
43	812	95.8	837	24	AAD29353	Escherichia coli s
44	812	95.8	837	24	AAD29370	Escherichia coli s
45	807	95.2	888	16	AAQ93069	FimH protein gene

ALIGNMENTS

RESULT 1
AAD29358
ID AAD29358 standard; DNA; 837 BP.
XX
AC AAD29358;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain B240 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli B240.
XX
FH Key Location/Qualifiers
FT CDS 1..837
FT /*tag= a

```
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000;. 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18419.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B240 FimH DNA.
XX
SQ   Sequence 837 BP; 193 A; 200 C; 220 G; 224 T; 0 other;

Alignment Scores:
Pred. No.:              4.05e-91        Length:         837
Score:                  848.00          Matches:        161
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     24              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29358 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGTTGGTGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCGTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 2
AAD29368
ID   AAD29368 standard; DNA; 837 BP.
XX
AC   AAD29368;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC80 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli EC80.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag=  a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:79..81, aa:Ala)
FT                   /transl_except= (pos:493..495, aa:Val)
FT                   /transl_except= (pos:511..513, aa:Ile)
FT                   /transl_except= (pos:805..807, aa:Lys)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18429.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC80 FimH DNA.
XX
SQ   Sequence 837 BP; 191 A; 200 C; 223 G; 223 T; 0 other;

Alignment Scores:
```

Pred. No.:	2.75e-90	Length:	837
Score:	841.00	Matches:	159
Percent Similarity:	99.38%	Conservative:	1
Best Local Similarity:	98.76%	Mismatches:	1
Query Match:	99.17%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29368 (1-837)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCGTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 3
AAQ93062
ID AAQ93062 standard; DNA; 900 BP.
XX
AC AAQ93062;
XX
DT 15-MAR-1996 (first entry)
XX
DE FimH protein gene from E. coli #1.
XX
KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FimA; FimF; FimG; receptor binding site; ss.
XX
OS Escherichia coli.
XX

FH	Key	Location/Qualifiers
FT	sig_peptide	1..63
FT		/*tag= a
FT	mat_peptide	64..900
FT		/*tag= b

XX
PN WO9520657-A1.
XX
PD 03-AUG-1995.
XX
PF 27-JAN-1995; 95WO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR WPI; 1995-275442/36.
XX
PT Receptor specific bacterial adhesins - useful for targetting active
PT compounds and microbial cells to locations of receptors
XX
PS Example 5; Page 114; 152pp; English.
XX
CC The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC coli clinical isolates. FimH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FimH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimA and
CC the minor components FimF and FimG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of variant FimH adhesins which
CC may be useful for targetting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.
XX
SQ Sequence 900 BP; 204 A; 212 C; 241 G; 243 T; 0 other;

Alignment Scores:

Pred. No.:	3.05e-90	Length:	900
Score:	841.00	Matches:	159
Percent Similarity:	99.38%	Conservative:	1
Best Local Similarity:	98.76%	Mismatches:	1
Query Match:	99.17%	Indels:	0
DB:	16	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93062 (1-900)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    139 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db    379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
```

```
Db    559 CGTGATGTCACCGTTACTCTGCCGGACTACCGTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy    161 Tyr 161
          |||
Db    619 TAT 621

RESULT 4
AAD29357
ID   AAD29357 standard; DNA; 837 BP.
XX
AC   AAD29357;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B238 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli B238.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag=  a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:70..72, aa:Leu)
FT                   /transl_except= (pos:76..78, aa:Ile)
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /transl_except= (pos:817..819, aa:Gly)
FT                   /transl_except= (pos:820..822, aa:Ala)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18418.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B238 FimH DNA.
XX
SQ   Sequence 837 BP; 192 A; 202 C; 223 G; 220 T; 0 other;

Alignment Scores:
Pred. No.:              1.87e-89        Length:         837
Score:                  834.00          Matches:        158
Percent Similarity:     98.76%          Conservative:   1
Best Local Similarity:  98.14%          Mismatches:     2
Query Match:            98.35%          Indels:         0
DB:                     24              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29357 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTACCCGGAAACCATTACAGATTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGTGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGTGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 5
AAD29372
ID   AAD29372 standard; DNA; 837 BP.
XX
AC   AAD29372;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain J96 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli J96.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag=  a
FT                   /product= "E. coli FimH protein"
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
```

XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR P-PSDB; AAE18433.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Example 1; Fig 1; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain J96 FimH DNA.
XX
SQ Sequence 837 BP; 191 A; 201 C; 222 G; 223 T; 0 other;

Alignment Scores:

Pred. No.:	1.87e-89	Length:	837
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29372 (1-837)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

RESULT 6
AAD29361

ID AAD29361 standard; DNA; 840 BP.
XX
AC AAD29361;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC42 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli EC42.
XX
FH Key Location/Qualifiers
FT CDS 1..840
FT /*tag= a
FT /product= "E. coli FimH protein"
FT /note= "CDS does not include start codon"
FT /partial
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR P-PSDB; AAE18422.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Example 1; Fig 1; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC42 FimH DNA.
XX
SQ Sequence 840 BP; 193 A; 202 C; 224 G; 221 T; 0 other;

Alignment Scores:

Pred. No.:	1.88e-89	Length:	840
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29361 (1-840)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGATTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGTGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGTGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 7
AAD29363
ID   AAD29363 standard; DNA; 840 BP.
XX
AC   AAD29363;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC56 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli EC56.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..840
FT                   /*tag=  a
FT                   /product= "E. coli FimH protein"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18424.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC56 FimH DNA.
XX
SQ   Sequence 840 BP; 193 A; 201 C; 222 G; 224 T; 0 other;

Alignment Scores:
Pred. No.:                1.88e-89        Length:          840
Score:                    834.00          Matches:         158
Percent Similarity:       98.76%          Conservative:    1
Best Local Similarity:    98.14%          Mismatches:      2
Query Match:              98.35%          Indels:          0
DB:                       24              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29363 (1-840)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 8
AAQ93070
ID   AAQ93070 standard; DNA; 900 BP.
XX
AC   AAQ93070;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli #2.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
```

```
FT                   /*tag=  a
FT   mat_peptide     64..900
FT                   /*tag=  b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 205 A; 213 C; 240 G; 242 T; 0 other;

Alignment Scores:
Pred. No.:                2.07e-89        Length:          900
Score:                    834.00          Matches:         158
Percent Similarity:       98.76%          Conservative:    1
Best Local Similarity:    98.14%          Mismatches:      2
Query Match:              98.35%          Indels:          0
DB:                       16              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93070 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621

RESULT 9
AAQ93071
ID   AAQ93071 standard; DNA; 900 BP.
XX
AC   AAQ93071;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli clinical isolate CI#3.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
OS   Escherichia coli clinical isolate CI#3.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
FT                   /*tag=  a
FT   mat_peptide     64..900
FT                   /*tag=  b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   P-PSDB; AAR76769.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
```

```
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 205 A; 214 C; 241 G; 240 T; 0 other;

Alignment Scores:
Pred. No.:                2.07e-89        Length:          900
Score:                    834.00          Matches:         158
Percent Similarity:       98.76%          Conservative:    1
Best Local Similarity:    98.14%          Mismatches:      2
Query Match:              98.35%          Indels:          0
DB:                       16              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93071 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTACCCGGAAACCATTACAGATTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGTGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGTGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621

RESULT 10
AAA66015
ID   AAA66015 standard; DNA; 903 BP.
XX
AC   AAA66015;
XX
DT   05-OCT-2000  (first entry)
XX
DE   E. coli proliferation associated coding sequence SEQ ID NO:207.
XX
KW   Escherichia coli; E. coli; proliferation; inhibition; screening;
KW   antimicrobial; bacterial growth; antisense therapy; antibacterial; ds.
XX
OS   Escherichia coli.
XX
PN   WO200044906-A2.
XX
PD   03-AUG-2000.
XX
PF   27-JAN-2000; 2000WO-US02200.
XX
PR   27-JAN-1999;   99US-0117405.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Zyskind J,  Ohlsen KL,  Trawick J,  Forsyth RA,  Froelich JM,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2000-514822/46.
DR   P-PSDB; AAB16009.
XX
PT   Novel polynucleotides and polypeptides associated with microorganism
PT   proliferation, used to identify inhibitors of bacterial growth and
PT   proliferation, for use in antisense therapy -
XX
PS   Claim 8; Page 162; 316pp; English.
XX
CC   AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC   sequences derived from Escherichia coli which inhibit E. coli
CC   proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC   nucleotide and protein sequences associated with E. coli proliferation.
CC   AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC   proliferation inhibiting nucleotide inserts in an example from the
CC   present invention. Methods from the present invention can be used to
CC   identify a proliferation- required gene in a microorganism, by contacting
CC   a microorganism with a proliferation-required gene activity inhibitory
CC   nucleic acid identified in another organism, and determining if
CC   inhibition occurs in the second microorganism. The nucleic acid sequences
CC   identified as being required for bacterial growth and proliferation, can
CC   be used for antisense therapy for killing bacteria.
XX
SQ   Sequence 903 BP; 206 A; 213 C; 240 G; 244 T; 0 other;

Alignment Scores:
Pred. No.:                2.08e-89        Length:          903
Score:                    834.00          Matches:         158
Percent Similarity:       98.76%          Conservative:    1
Best Local Similarity:    98.14%          Mismatches:      2
Query Match:              98.35%          Indels:          0
DB:                       21              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAA66015 (1-903)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
```

```
                 ||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621
```

RESULT 11
ABK11187
ID ABK11187 standard; DNA; 903 BP.
XX
AC ABK11187;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding Escherichia coli type I pilin protein, FimH.
XX
KW Immune response; primate; immunoglobulin; urogenital tract infection;
KW FimC-FimH complex; FimCH; IgG; human; urinary tract infection; UTI;
KW bladder infection; kidney infection; Enterobacteriaceae; bactourea;
KW pregnant woman; diabetic; immunocompromised; HIV; cancer;
KW human immunodeficiency virus infection; end stage renal disease;
KW type I pilin protein; adhesin; FimH; ds.
XX
OS Escherichia coli isolate J96.
XX
FH Key   Location/Qualifiers
FT CDS   1..903
FT   /*tag= a
FT   /product= "FimH"
XX
PN WO200215928-A1.
XX
PD 28-FEB-2002.
XX
PF 28-NOV-2000; 2000WO-US32398.
XX
PR 18-AUG-2000; 2000US-226146P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Ballou WR;
XX
DR WPI; 2002-280859/32.
DR P-PSDB; AAU77488.
XX
PT Stimulating immune response in a primate for preventing, treating
PT bacterial induced diseases such as diseases of urinary tract, by
PT administering bacterial adhesive proteins, preferably FimC-FimH
PT polypeptide complex -
XX
PS Disclosure; Page 88-89; 92pp; English.
XX
CC The present invention relates to a method of inducing an immune response
CC in a primate. The response involves immunoglobulin (Ig) molecules that
CC bind a bacterial adhesin protein, preferably an attachment domain of
CC a type 1 pilin polypeptide (e.g. FimH) associated with a bacterium
CC causing urogenital tract infections (e.g. Escherichia coli). The method
CC comprises administering a purified FimH polypeptide, a FimC-FimH (FimCH)
CC complex, or immunogenic fragments of these. The method is useful for
CC inducing IgG molecules in a primate, especially human, to reduce or
CC prevent the incidence of urogenital tract infections, particularly
CC urinary tract infection (UTI), bladder infection, or kidney infection,
CC caused by a bacterium of the family Enterobacteriaceae, preferably
CC E. coli. The method can be used in a human subject that has suffered
CC more than two urogenital infections within one year, has asymptomatic
CC bactourea, is a pregnant woman or a diabetic, is immunocompromised, has
CC a human immunodeficiency virus (HIV) infection, has cancer, is in
CC remission from cancer, or is at risk for end stage renal disease. The
CC method is useful for vaccinating a primate against urogenital tract
CC infections, for treating or ameliorating the symptoms of urogenital
CC tract infections, and also for slowing or preventing progression of
CC a urinary tract infection into end stage renal disease. The present
CC sequence encodes E. coli FimH protein.
XX
SQ Sequence 903 BP; 206 A; 213 C; 240 G; 244 T; 0 other;

Alignment Scores:

Pred. No.:	2.08e-89	Length:	903
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x ABK11187 (1-903)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621
```

RESULT 12
AAD29376
ID AAD29376 standard; DNA; 903 BP.
XX
AC AAD29376;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain J96 FimH DNA plus signal sequence.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli J96.
XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX

PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Example 2; Page 96; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain J96 FimH DNA plus signal
CC sequence, used in the exemplification of the invention.
XX
SQ Sequence 903 BP; 206 A; 213 C; 240 G; 244 T; 0 other;

Alignment Scores:

Pred. No.:	2.08e-89	Length:	903
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29376 (1-903)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621
```

RESULT 13

ABK23921
ID ABK23921 standard; DNA; 4637 BP.
XX
AC ABK23921;
XX
DT 09-APR-2002 (first entry)
XX
DE Vector pFIMAFGH expressing pilin genes fimF, G and H.
XX
KW Vaccine; molecular scaffold; pilus; pilin; HBcAg; antigen;
KW hepatitis B virus capsid protein; JUN; FOS; HIV gp140;
KW measles virus N protein; bee venom phospholipase; Th type 2 T-helper;
KW Th2; Sinbis virus E2 protein; amyloid beta; influenza M2 antigen;
KW human immunodeficiency virus infection; viral hepatitis; measles;
KW chicken pox; pneumonia; tuberculosis; syphilis; malaria; allergy;
KW cancer; chronic disease; arthritis; colitis; diabetes;
KW multiple sclerosis; ds.
XX
OS Escherichia coli.
XX
PN WO200185208-A2.
XX
PD 15-NOV-2001.
XX
PF 02-MAY-2001; 2001WO-IB00741.
XX
PR 05-MAY-2000; 2000US-202341P.
XX
PA (CYTO-) CYTOS BIOTECHNOLOGY AG.
PA (SEBB/) SEBBEL P.
PA (DUNA/) DUNANT N.
PA (BACH/) BACHMANN M.
PA (TISS/) TISSOT A.
PA (LECH/) LECHENER F.
XX
PI Sebbel P, Dunant N, Bachmann M, Tissot A, Lechener F;
XX
DR WPI; 2002-055561/07.
XX
PT New composition, useful for vaccine production, comprises antigen or
PT antigenic determinant and non-natural molecular scaffold comprising
PT organizer and core particle such as bacterial pilus or pilin protein
PT -s
XX
PS Example 34; Page 265-267; 287pp; English.
XX
CC The invention relates to a composition comprising: (a) a non-natural
CC molecular scaffold (molecular scaffold) which comprises a core
CC particle such as a bacterial pilus or pilin protein, a recombinant form
CC of the protein, a virus-like particle or a hepatitis B virus capsid
CC protein (HBcAg), and an organiser; and (b) an antigen or antigenic
CC determinant, where the molecular scaffold and antigenic determinant
CC interact to form an ordered and repetitive antigen array. Suitable
CC antigenic determinants include JUN, FOS, HIV gp140, measles virus N
CC protein, bee venom phospholipase, Sinbis virus E2 protein, amyloid beta
CC derived peptides and influenza M2 antigen. The composition (or vaccine)
CC is useful for immunisation, by administration to a subject, where the
CC administration produces an immune response, such as humoral, cellular or
CC protective immune response, preferably a Th type 2 T-helper (Th2)
CC response that is specific for the antigenic determinant. The
CC administration induces antibodies specific for the antigenic determinant
CC of a subtype corresponding to the Th2 subtype in the subject. The subject
CC does not generate a Th2 subtype that is specific for pilus or pilin
CC polypeptide or antigenic determinant. The composition is useful for the
CC production of vaccines for prevention of infectious diseases such as
CC human immunodeficiency virus, viral hepatitis, measles, chicken pox,
CC pneumonia, tuberculosis, syphilis, malaria, and for treating allergy,
CC cancer, and chronic diseases induced or accelerated by a Th1 type
CC immune response, such as arthritis, colitis, diabetes and multiple
CC sclerosis. The composition is useful to generate defined self-specific
CC antibodies and specific immune responses of the Th2 type and allows the
CC creation of highly efficient vaccines against infectious diseases, and
CC for treating allergy, cancer, and chronic diseases induced or accelerated

```
CC   by a Th1 type immune response. The present sequent is a nucleic acid
CC   encoding a composition of the invention or a component of one.
XX
SQ   Sequence 4637 BP; 1124 A; 1156 C; 1188 G; 1169 T; 0 other;

Alignment Scores:
Pred. No.:              2.08e-88        Length:         4637
Score:                  834.00          Matches:        158
Percent Similarity:     98.76%          Conservative:   1
Best Local Similarity:  98.14%          Mismatches:     2
Query Match:            98.35%          Indels:         0
DB:                     24              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x ABK23921 (1-4637)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3441 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 3500

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3501 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 3560

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3561 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 3620

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3621 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 3680

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db    3681 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 3740

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3741 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 3800

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3801 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 3860

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    3861 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 3920

Qy     161 Tyr 161
           |||
Db    3921 TAT 3923

RESULT 14
AAD29374
ID   AAD29374 standard; DNA; 7416 BP.
XX
AC   AAD29374;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Plasmid pCGA126-1 for construction of plasmid pGA139-1.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; plasmid pCGA126-1;
KW   ds.
XX
OS   Chimeric - Escherichia coli.
OS   Chimeric - Unidentified.
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Example 2; Page 93-95; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is plasmid pCGA126-1 used for constructing plasmid
CC   pGA139-1 used in the exemplification of the invention. Plasmid pCGA126-1
CC   consists of Escherichia coli FimH DNA, FimC and its native signal and
CC   lac p/o.
XX
SQ   Sequence 7416 BP; 1858 A; 1807 C; 1908 G; 1843 T; 0 other;

Alignment Scores:
Pred. No.:              4.03e-88        Length:         7416
Score:                  834.00          Matches:        158
Percent Similarity:     98.76%          Conservative:   1
Best Local Similarity:  98.14%          Mismatches:     2
Query Match:            98.35%          Indels:         0
DB:                     24              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29374 (1-7416)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1244 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 1303

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1304 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 1363

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 1423

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1424 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 1483

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db    1484 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 1543

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1544 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 1603

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1604 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 1663

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    1664 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 1723
```

```
Qy     161 Tyr 161
           |||
Db    1724 TAT 1726

RESULT 15
ABK23922
ID   ABK23922 standard; DNA; 9299 BP.
XX
AC   ABK23922;
XX
DT   09-APR-2002  (first entry)
XX
DE   Vector pFIMAICFGH expressing pilin genes fimA, I, C, F, G and H.
XX
KW   Vaccine; molecular scaffold; pilus; pilin; HBcAg; antigen;
KW   hepatitis B virus capsid protein; JUN; FOS; HIV gp140;
KW   measles virus N protein; bee venom phospholipase; Th type 2 T-helper;
KW   Th2; Sinbis virus E2 protein; amyloid beta; influenza M2 antigen;
KW   human immunodeficiency virus infection; viral hepatitis; measles;
KW   chicken pox; pneumonia; tuberculosis; syphilis; malaria; allergy;
KW   cancer; chronic disease; arthritis; colitis; diabetes;
KW   multiple sclerosis; ds.
XX
OS   Escherichia coli.
XX
PN   WO200185208-A2.
XX
PD   15-NOV-2001.
XX
PF   02-MAY-2001; 2001WO-IB00741.
XX
PR   05-MAY-2000; 2000US-202341P.
XX
PA   (CYTO-) CYTOS BIOTECHNOLOGY AG.
PA   (SEBB/) SEBBEL P.
PA   (DUNA/) DUNANT N.
PA   (BACH/) BACHMANN M.
PA   (TISS/) TISSOT A.
PA   (LECH/) LECHENER F.
XX
PI   Sebbel P,  Dunant N,  Bachmann M,  Tissot A,  Lechener F;
XX
DR   WPI; 2002-055561/07.
XX
PT   New composition, useful for vaccine production, comprises antigen or
PT   antigenic determinant and non-natural molecular scaffold comprising
PT   organizer and core particle such as bacterial pilus or pilin protein
PT    -s
XX
PS   Example 34; Page 268-273; 287pp; English.
XX
CC   The invention relates to a composition comprising: (a) a non-natural
CC   molecular scaffold (molecular scaffold) which comprises a core
CC   particle such as a bacterial pilus or pilin protein, a recombinant form
CC   of the protein, a virus-like particle or a hepatitis B virus capsid
CC   protein (HBcAg), and an organiser; and (b) an antigen or antigenic
CC   determinant, where the molecular scaffold and antigenic determinant
CC   interact to form an ordered and repetitive antigen array. Suitable
CC   antigenic determinants include JUN, FOS, HIV gp140, measles virus N
CC   protein, bee venom phospholipase, Sinbis virus E2 protein, amyloid beta
CC   derived peptides and influenza M2 antigen. The composition (or vaccine)
CC   is useful for immunisation, by administration to a subject, where the
CC   administration produces an immune response, such as humoral, cellular or
CC   protective immune response, preferably a Th type 2 T-helper (Th2)
CC   response that is specific for the antigenic determinant. The
CC   administration induces antibodies specific for the antigenic determinant
CC   of a subtype corresponding to the Th2 subtype in the subject. The subject
CC   does not generate a Th2 subtype that is specific for pilus or pilin
CC   polypeptide or antigenic determinant. The composition is useful for the
CC   production of vaccines for prevention of infectious diseases such as
CC   human immunodeficiency virus, viral hepatitis, measles, chicken pox,
CC   pneumonia, tuberculosis, syphilis, malaria, and for treating allergy,
CC   cancer, and chronic diseases induced or accelerated by a Th1 type
CC   immune response, such as arthritis, colitis, diabetes and multiple
CC   sclerosis. The composition is useful to generate defined self-specific
CC   antibodies and specific immune responses of the Th2 type and allows the
CC   creation of highly efficient vaccines against infectious diseases, and
CC   for treating allergy, cancer, and chronic diseases induced or accelerated
CC   by a Th1 type immune response. The present sequent is a nucleic acid
CC   encoding a composition of the invention or a component of one.
XX
SQ   Sequence 9299 BP; 2371 A; 2185 C; 2410 G; 2333 T; 0 other;

Alignment Scores:
Pred. No.:                5.54e-88      Length:          9299
Score:                    834.00        Matches:         158
Percent Similarity:       98.76%        Conservative:    1
Best Local Similarity:    98.14%        Mismatches:      2
Query Match:              98.35%        Indels:          0
DB:                       24            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x ABK23922 (1-9299)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8109 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 8168

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8169 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 8228

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8229 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 8288

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8289 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 8348

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db    8349 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 8408

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8409 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 8468

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8469 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 8528

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    8529 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 8588

Qy     161 Tyr 161
           |||
Db    8589 TAT 8591

RESULT 16
AAD29382
ID   AAD29382 standard; DNA; 837 BP.
XX
AC   AAD29382;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain G162 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli G162.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
```

```
FT                   /*tag= a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18435.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Example 1; Page 98; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain G162 FimH DNA.
XX
SQ   Sequence 837 BP; 192 A; 201 C; 221 G; 223 T; 0 other;

Alignment Scores:
Pred. No.:                4.25e-89        Length:         837
Score:                    831.00          Matches:        157
Percent Similarity:       98.76%          Conservative:   2
Best Local Similarity:    97.52%          Mismatches:     2
Query Match:              98.00%          Indels:         0
DB:                       24              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29382 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 17
AAD29354
ID   AAD29354 standard; DNA; 837 BP.
XX
AC   AAD29354;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B217 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli B217.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag= a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:184..186, aa:Ala)
FT                   /transl_except= (pos:208..210, aa:Ser)
FT                   /transl_except= (pos:232..234, aa:Asn)
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /transl_except= (pos:700..702, aa:Ala)
FT                   /transl_except= (pos:730..732, aa:Gly)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18415.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
```

```
CC   The present sequence is Escherichia coli strain B217 FimH DNA.
XX
SQ   Sequence 837 BP; 193 A; 201 C; 221 G; 222 T; 0 other;

Alignment Scores:
Pred. No.:                5.58e-89      Length:          837
Score:                    830.00        Matches:         157
Percent Similarity:       98.14%        Conservative:    1
Best Local Similarity:    97.52%        Mismatches:      3
Query Match:              97.88%        Indels:          0
DB:                       24            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29354 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 18
AAQ93074
ID   AAQ93074 standard; DNA; 900 BP.
XX
AC   AAQ93074;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli clinical isolate U221-3.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
OS   Escherichia coli clinical isolate U221-3.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
FT                   /*tag=  a
FT   mat_peptide     64..900
FT                   /*tag=  b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;    95WO-DK00042.
XX
PR   27-JAN-1994;    94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   P-PSDB; AAR76772.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 205 A; 214 C; 240 G; 241 T; 0 other;

Alignment Scores:
Pred. No.:                6.18e-89      Length:          900
Score:                    830.00        Matches:         157
Percent Similarity:       98.14%        Conservative:    1
Best Local Similarity:    97.52%        Mismatches:      3
Query Match:              97.88%        Indels:          0
DB:                       16            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93074 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498
```

```
Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    559 CGTGATGTCACCGTTACTCTGCCGGACTACCCAGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy    161 Tyr 161
          |||
Db    619 TAT 621

RESULT 19
AAD29366
ID   AAD29366 standard; DNA; 837 BP.
XX
AC   AAD29366;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC61 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli EC61.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag= a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:7..9, aa:Cys)
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18427.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC61 FimH DNA.
XX
SQ   Sequence 837 BP; 191 A; 202 C; 221 G; 223 T; 0 other;

Alignment Scores:

Pred. No.:                1.67e-88      Length:          837
Score:                    826.00        Matches:         156
Percent Similarity:       98.14%        Conservative:    2
Best Local Similarity:    96.89%        Mismatches:      3
Query Match:              97.41%        Indels:          0
DB:                       24            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29366 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAGAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 20
AAD29356
ID   AAD29356 standard; DNA; 840 BP.
XX
AC   AAD29356;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B228 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli B228.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..840
FT                   /*tag= a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
```

PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR P-PSDB; AAE18417.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection
XX
PS Example 1; Page 60; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain B228 FimH DNA.
XX
SQ Sequence 840 BP; 196 A; 197 C; 222 G; 225 T; 0 other;

Alignment Scores:

Pred. No.:	2.2e-88	Length:	840
Score:	825.00	Matches:	156
Percent Similarity:	97.52%	Conservative:	1
Best Local Similarity:	96.89%	Mismatches:	4
Query Match:	97.29%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29356 (1-840)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db     496 CATGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
```

```
           |||
Db     556 TAT 558
```

RESULT 21
AAQ93067
ID AAQ93067 standard; DNA; 900 BP.
XX
AC AAQ93067;
XX
DT 15-MAR-1996 (first entry)
XX
DE FimH protein gene from E. coli clinical isolate MJ#9-3.
XX
KW FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW FimA; FimF; FimG; receptor binding site; ss.
XX
OS Escherichia coli clinical isolate MJ#9-3.
XX

FH	Key	Location/Qualifiers
FT	sig_peptide	1..63
FT		/*tag= a
FT	mat_peptide	64..900
FT		/*tag= b

XX
PN WO9520657-A1.
XX
PD 03-AUG-1995.
XX
PF 27-JAN-1995; 95WO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR WPI; 1995-275442/36.
DR P-PSDB; AAR76773.
XX
PT Receptor specific bacterial adhesins - useful for targetting active
PT compounds and microbial cells to locations of receptors
XX
PS Example 5; Page 69-74; 152pp; English.
XX
CC The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC coli clinical isolates. FimH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FimH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimA and
CC the minor components FimF and FimG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of variant FimH adhesins which
CC may be useful for targetting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.
XX
SQ Sequence 900 BP; 207 A; 210 C; 240 G; 243 T; 0 other;

Alignment Scores:

Pred. No.:	2.43e-88	Length:	900
Score:	825.00	Matches:	156
Percent Similarity:	97.52%	Conservative:	1
Best Local Similarity:	96.89%	Mismatches:	4
Query Match:	97.29%	Indels:	0
DB:	16	Gaps:	0

```
US-09-900-575-29_COPY_26_186 (1-161) x AAQ93067 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTCCCG 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
             ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     559 CATGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621

RESULT 22
AAQ93073
ID   AAQ93073 standard; DNA; 900 BP.
XX
AC   AAQ93073;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli clinical isolate KS54.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
OS   Escherichia coli clinical isolate KS54.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
FT                   /*tag=  a
FT   mat_peptide     64..900
FT                   /*tag=  b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   P-PSDB; AAR76771.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 207 A; 210 C; 240 G; 243 T; 0 other;

Alignment Scores:
Pred. No.:                2.43e-88      Length:          900
Score:                    825.00        Matches:         156
Percent Similarity:       97.52%        Conservative:    1
Best Local Similarity:    96.89%        Mismatches:      4
Query Match:              97.29%        Indels:          0
DB:                       16            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93073 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
             ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     559 CATGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
```

```
Db    619 TAT 621

RESULT 23
AAQ93068
ID   AAQ93068 standard; DNA; 900 BP.
XX
AC   AAQ93068;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli clinical isolate MJ#31-3.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
OS   Escherichia coli clinical isolate MJ#31-3.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
FT                   /*tag=  a
FT   mat_peptide     64..900
FT                   /*tag=  b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   P-PSDB; AAR76774.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 207 A; 209 C; 240 G; 244 T; 0 other;

Alignment Scores:
Pred. No.:                2.43e-88        Length:          900
Score:                    825.00          Matches:         156
Percent Similarity:       97.52%          Conservative:    1
Best Local Similarity:    96.89%          Mismatches:      4
Query Match:              97.29%          Indels:          0
DB:                       16              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93068 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db     559 CATGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621

RESULT 24
AAD29371
ID   AAD29371 standard; DNA; 837 BP.
XX
AC   AAD29371;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain G189 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli G189.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag=  a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
```

```
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18432.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Example 1; Page 66; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain G189 FimH DNA.
XX
SQ   Sequence 837 BP; 196 A; 200 C; 220 G; 221 T; 0 other;

Alignment Scores:
Pred. No.:                2.88e-88      Length:          837
Score:                    824.00        Matches:         156
Percent Similarity:       97.52%        Conservative:    1
Best Local Similarity:    96.89%        Mismatches:      4
Query Match:              97.17%        Indels:          0
DB:                       24            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29371 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAAAAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTAGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 25
AAD29510
ID   AAD29510 standard; DNA; 837 BP.
XX
AC   AAD29510;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC189 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   vaccine; urinary tract infection; bladder; therapy; ds.
XX
OS   Escherichia coli EC189.
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Disclosure; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC189 FimH DNA.
XX
SQ   Sequence 837 BP; 196 A; 200 C; 220 G; 221 T; 0 other;

Alignment Scores:
Pred. No.:                2.88e-88      Length:          837
Score:                    824.00        Matches:         156
Percent Similarity:       97.52%        Conservative:    1
Best Local Similarity:    96.89%        Mismatches:      4
Query Match:              97.17%        Indels:          0
DB:                       24            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29510 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375
```

```
Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAAAAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTAGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 26
AAD29367
ID AAD29367 standard; DNA; 837 BP.
XX
AC AAD29367;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC62 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli EC62.
XX

FH	Key	Location/Qualifiers
FT	CDS	1..837
FT		/*tag= a
FT		/product= "E. coli FimH protein"
FT		/transl_except= (pos:526..528, aa:Arg)
FT		/transl_except= (pos:601..603, aa:His)
FT		/note= "CDS does not include start and stop codon"
FT		/partial

XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR P-PSDB; AAE18428.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Example 1; Fig 1; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC62 FimH DNA.
XX

SQ Sequence 837 BP; 195 A; 202 C; 221 G; 219 T; 0 other;

Alignment Scores:			
Pred. No.:	4.99e-88	Length:	837
Score:	822.00	Matches:	155
Percent Similarity:	98.14%	Conservative:	3
Best Local Similarity:	96.27%	Mismatches:	3
Query Match:	96.93%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29367 (1-837)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTCATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAGGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATGGCTGTGCTAATTTTGCGACAGACCAATAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCAGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 27
AAD29369
ID AAD29369 standard; DNA; 837 BP.
XX
AC AAD29369;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC89 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli EC89.
XX

FH	Key	Location/Qualifiers
FT	CDS	1..837
FT		/*tag= a
FT		/product= "E. coli FimH protein"
FT		/transl_except= (pos:526..528, aa:Arg)
FT		/transl_except= (pos:601..603, aa:His)
FT		/transl_except= (pos:676..678, aa:Thr)
FT		/transl_except= (pos:679..681, aa:Ala)
FT		/transl_except= (pos:691..693, aa:Ile)
FT		/transl_except= (pos:694..696, aa:Val)

```
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18430.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Example 1; Page 65; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC89 FimH DNA.
XX
SQ   Sequence 837 BP; 188 A; 202 C; 226 G; 221 T; 0 other;

Alignment Scores:
Pred. No.:                4.99e-88      Length:          837
Score:                    822.00        Matches:         155
Percent Similarity:       98.14%        Conservative:    3
Best Local Similarity:    96.27%        Mismatches:      3
Query Match:              96.93%        Indels:          0
DB:                       24            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29369 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||||:::||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 28
AAQ93063
ID   AAQ93063 standard; DNA; 900 BP.
XX
AC   AAQ93063;
XX.
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli clinical isolate F-18.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
OS   Escherichia coli clinical isolate F-18.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
FT                   /*tag=  a
FT   mat_peptide     64..900
FT                   /*tag=  b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   P-PSDB; AAR76770.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 206 A; 214 C; 241 G; 239 T; 0 other;
```

```
Alignment Scores:
Pred. No.:                5.52e-88        Length:          900
Score:                    822.00          Matches:         155
Percent Similarity:       98.14%          Conservative:    3
Best Local Similarity:    96.27%          Mismatches:      3
Query Match:              96.93%          Indels:          0
DB:                       16              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93063 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::||||||||||||||||||:::||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621

RESULT 29
AAD29359
ID   AAD29359 standard; DNA; 837 BP.
XX
AC   AAD29359;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B242 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli B242.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag=  a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /transl_except= (pos:835..837, aa:Glx)
FT                   /note= "There is an apparent deletion of 3 codons which
FT                   alters the reading frame; CDS does not include start and
FT                   stop codon"
FT                   /partial
XX

PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18420.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Example 1; Fig 2; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B242 FimH protein.
XX
SQ   Sequence 837 BP; 192 A; 200 C; 225 G; 220 T; 0 other;

Alignment Scores:
Pred. No.:                8.62e-88        Length:          837
Score:                    820.00          Matches:         155
Percent Similarity:       96.89%          Conservative:    1
Best Local Similarity:    96.27%          Mismatches:      5
Query Match:              96.70%          Indels:          0
DB:                       24              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29359 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db      76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAACCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
```

```
Db    496 CATGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 30
AAD29362
ID AAD29362 standard; DNA; 837 BP.
XX
AC AAD29362;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC45 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli EC45.
XX

FH	Key	Location/Qualifiers
FT	CDS	1..837
FT		/*tag= a
FT		/product= "E. coli FimH protein"
FT		/note= "CDS does not include start and stop codon"
FT		/partial

XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR P-PSDB; AAE18423.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Example 1; Fig 1; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC45 FimH DNA.
XX
SQ Sequence 837 BP; 193 A; 202 C; 224 G; 218 T; 0 other;

Alignment Scores:

Pred. No.:	1.13e-87	Length:	837
Score:	819.00	Matches:	154
Percent Similarity:	98.14%	Conservative:	4
Best Local Similarity:	95.65%	Mismatches:	3
Query Match:	96.58%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29362 (1-837)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||:::|||||||||||||||||||||||:::|||||||||||||||||||
Db    196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 31
AAD29373
ID AAD29373 standard; DNA; 837 BP.
XX
AC AAD29373;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain NU14 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli NU14.
XX

FH	Key	Location/Qualifiers
FT	CDS	1..837
FT		/*tag= a
FT		/product= "E. coli FimH protein"
FT		/transl_except= (pos:526..528, aa:Arg)
FT		/transl_except= (pos:601..603, aa:His)
FT		/note= "CDS does not include start and stop codon"
FT		/partial

XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.

```
DR   P-PSDB; AAE18434.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Example 1; Page 67; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain NU14 FimH DNA.
XX
SQ   Sequence 837 BP; 193 A; 202 C; 224 G; 218 T; 0 other;

Alignment Scores:
Pred. No.:              1.13e-87        Length:         837
Score:                  819.00          Matches:        154
Percent Similarity:     98.14%          Conservative:   4
Best Local Similarity:  95.65%          Mismatches:     3
Query Match:            96.58%          Indels:         0
DB:                     24              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29373 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||:::|||||||||||||||||:::|||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 32
AAD29509
ID   AAD29509 standard; DNA; 837 BP.
XX
AC   AAD29509;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B203 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   vaccine; urinary tract infection; bladder; therapy; ds.
XX
OS   Escherichia coli B203.
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection  -
XX
PS   Disclosure; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B203 FimH DNA.
XX
SQ   Sequence 837 BP; 193 A; 201 C; 225 G; 218 T; 0 other;

Alignment Scores:
Pred. No.:              1.13e-87        Length:         837
Score:                  819.00          Matches:        154
Percent Similarity:     98.14%          Conservative:   4
Best Local Similarity:  95.65%          Mismatches:     3
Query Match:            96.58%          Indels:         0
DB:                     24              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29509 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||:::|||||||||||||||||:::|||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
```

```
          ||||||
```

```
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18421.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain DS17 FimH DNA.
XX
SQ   Sequence 837 BP; 194 A; 200 C; 223 G; 220 T; 0 other;

Alignment Scores:
Pred. No.:                1.49e-87      Length:          837
Score:                    818.00        Matches:         154
Percent Similarity:       97.52%        Conservative:    3
Best Local Similarity:    95.65%        Mismatches:      4
Query Match:              96.46%        Indels:          0
DB:                       24            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29360 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::||||||||||||||||||:::||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||:::  ||||||||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGCTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTTTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 35
AAD29352
ID   AAD29352 standard; DNA; 837 BP.
XX
AC   AAD29352;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B210 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli B210.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag= a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:421..423, aa:Arg)
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /transl_except= (pos:607..609, aa:Asp)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18413.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B210 FimH DNA.
XX
SQ   Sequence 837 BP; 186 A; 204 C; 228 G; 219 T; 0 other;

Alignment Scores:
Pred. No.:                1.96e-87      Length:          837
Score:                    817.00        Matches:         155
Percent Similarity:       96.89%        Conservative:    1
Best Local Similarity:    96.27%        Mismatches:      5
Query Match:              96.34%        Indels:          0
DB:                       24            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29352 (1-837)
```

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGATCGTAAAATATAGTGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCAGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGGTTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGCTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

```
RESULT 36
AAD29364
ID   AAD29364 standard; DNA; 837 BP.
XX
AC   AAD29364;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC58 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli EC58.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag= a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18425.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC58 FimH DNA.
XX
SQ   Sequence 837 BP; 195 A; 200 C; 222 G; 220 T; 0 other;
```

```
Alignment Scores:
Pred. No.:                2.58e-87        Length:          837
Score:                    816.00          Matches:         154
Percent Similarity:       97.52%          Conservative:    3
Best Local Similarity:    95.65%          Mismatches:      4
Query Match:              96.23%          Indels:          0
DB:                       24              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29364 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
              |||||||||:::|||||||||||||||||||||||||:::||||||||||||||||||||
Db     196 AGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

```
RESULT 37
AAD29365
ID   AAD29365 standard; DNA; 837 BP.
XX
AC   AAD29365;
```

```
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain EC60 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli EC60.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag=  a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:79..81, aa:Val)
FT                   /transl_except= (pos:196..198, aa:Gly)
FT                   /transl_except= (pos:208..210, aa:Asn)
FT                   /transl_except= (pos:232..234, aa:Ser)
FT                   /transl_except= (pos:355..357, aa:Val)
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /transl_except= (pos:805..807, aa:Arg)
FT                   /transl_except= (pos:817..819, aa:Ala)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18426.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain EC60 FimH DNA.
XX
SQ   Sequence 837 BP; 195 A; 200 C; 222 G; 220 T; 0 other;

Alignment Scores:
Pred. No.:                2.58e-87        Length:          837
Score:                    816.00          Matches:         154
Percent Similarity:       97.52%          Conservative:    3
Best Local Similarity:    95.65%          Mismatches:      4
Query Match:              96.23%          Indels:          0
DB:                       24              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29365 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           |||||||||:::|||||||||||||||||||||:::|||||||||||||||||||||||
Db     196 AGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 38
ABL52756
ID   ABL52756 standard; DNA; 1227 BP.
XX
AC   ABL52756;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Fusion protein prtY-FimH-prtY encoding sequence.
XX
KW   Exoproteinase; immunostimulant; vaccine; anchor peptide; gene;
KW   FimH; ciliated adhesive factor; ds.
XX
OS   Lactobacillus helveticus.
OS   Escherichia coli.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1227
FT                   /*tag=  a
FT                   /product= "PrtY-FimH-PrtY"
XX
PN   JP2002017357-A.
XX
PD   22-JAN-2002.
XX
PF   04-JUL-2000; 2000JP-0202442.
XX
PR   04-JUL-2000; 2000JP-0202442.
XX
PA   (CALV ) CALPIS SHOKUHIN KOGYO KK.
XX
DR   WPI; 2002-221706/28.
DR   P-PSDB; ABB09458.
XX
PT   New anchor peptide, useful for anchoring protein onto microbe -
XX
PS   Example 1; Page 8-10; 12pp; Japanese.
XX
CC   The invention relates to an anchor peptide derived from the
```

```
CC   exoproteinase of Lactobacillus helveticus FERM BP-6060.
CC   The activity of compositions of the invention may be described as
CC   immunostimulatory. The invention also includes a fusion protein,
CC   containing the anchor peptide and a signal peptide fused to a useful
CC   protein. The invention also includes a method for anchoring a useful
CC   protein onto the surface of a microbe. The microbe can be used as a
CC   vaccine. The current sequence represents an encoding sequence for a
CC   fusion protein referred to as prtY-FimH-prtY. This fusion protein
CC   contains a fragment of the FimH protein, which is a ciliated adhesive
CC   factor from E. coli, arranged between the the signal sequence and anchor
CC   sequences derived from L. helveticus exoproteinase (see ABB09456 and
CC   ABB09457).
XX
SQ   Sequence 1227 BP; 325 A; 285 C; 287 G; 330 T; 0 other;

Alignment Scores:
Pred. No.:                4.41e-87        Length:         1227
Score:                    816.00          Matches:        154
Percent Similarity:       97.52%          Conservative:   3
Best Local Similarity:    95.65%          Mismatches:     4
Query Match:              96.23%          Indels:         0
DB:                       24              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x ABL52756 (1-1227)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db     277 CCTGCCGTGAATGTGGGGCAAAAGCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 336

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     337 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 396

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||||:::||||||||||||||||||||||
Db     397 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 456

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     457 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 516

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     517 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 576

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     577 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 636

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     637 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 696

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     697 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 756

Qy     161 Tyr 161
           |||
Db     757 TAT 759

RESULT 39
AAQ93066
ID   AAQ93066 standard; DNA; 900 BP.
XX
AC   AAQ93066;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli clinical isolate MJ#2-2.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
OS   Escherichia coli clinical isolate MJ#2-2.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
FT                   /*tag= a
FT   mat_peptide     64..900
FT                   /*tag= b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   P-PSDB; AAR76776.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 208 A; 214 C; 240 G; 238 T; 0 other;

Alignment Scores:
Pred. No.:                3.75e-87        Length:         900
Score:                    815.00          Matches:        154
Percent Similarity:       97.52%          Conservative:   3
Best Local Similarity:    95.65%          Mismatches:     4
Query Match:              96.11%          Indels:         0
DB:                       16              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93066 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||:::|||||||||||||||||||:::||||||||||||||||||||||||
Db     259 GACGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 318
```

```
Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTCTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621
```

```
RESULT 40
AAQ93072
ID   AAQ93072 standard; DNA; 900 BP.
XX
AC   AAQ93072;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli clinical isolate CI#7.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
OS   Escherichia coli clinical isolate CI#7.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
FT                   /*tag=  a
FT   mat_peptide     64..900
FT                   /*tag=  b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;   95WO-DK00042.
XX
PR   27-JAN-1994;   94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   P-PSDB; AAR76766.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 206 A; 215 C; 239 G; 240 T; 0 other;
```

```
Alignment Scores:
Pred. No.:                3.75e-87      Length:          900
Score:                    815.00        Matches:         155
Percent Similarity:       97.52%        Conservative:    2
Best Local Similarity:    96.27%        Mismatches:      4
Query Match:              96.11%        Indels:          0
DB:                       16            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93072 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   |||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db     139 CCCGCCGTGAATGTGGGGCAACACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGCGAAACGCTGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621
```

```
RESULT 41
AAQ93065
ID   AAQ93065 standard; DNA; 900 BP.
XX
AC   AAQ93065;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli clinical isolate CI#4.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
```

```
OS   Escherichia coli clinical isolate CI#4.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
FT                   /*tag=  a
FT   mat_peptide     64..900
FT                   /*tag=  b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;    95WO-DK00042.
XX
PR   27-JAN-1994;    94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   P-PSDB; AAR76765.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 205 A; 210 C; 241 G; 244 T; 0 other;

Alignment Scores:
Pred. No.:                4.93e-87      Length:          900
Score:                    814.00        Matches:         154
Percent Similarity:       97.52%        Conservative:    3
Best Local Similarity:    95.65%        Mismatches:      4
Query Match:              95.99%        Indels:          0
DB:                       16            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93065 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::||||||   ||||||||||||:::|||||||||||||||||||||
Db     259 GGCGTGTTATCTAGTTTTTCCGAGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTTTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621

RESULT 42
AAQ93075
ID   AAQ93075 standard; DNA; 900 BP.
XX
AC   AAQ93075;
XX
DT   15-MAR-1996  (first entry)
XX
DE   FimH protein gene from E. coli clinical isolate CI#12.
XX
KW   FimH; type 1 fimbriae; organelle; adhesin; alpha-D-mannoside residue;
KW   FimA; FimF; FimG; receptor binding site; ss.
XX
OS   Escherichia coli clinical isolate CI#12.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..63
FT                   /*tag=  a
FT   mat_peptide     64..900
FT                   /*tag=  b
XX
PN   WO9520657-A1.
XX
PD   03-AUG-1995.
XX
PF   27-JAN-1995;    95WO-DK00042.
XX
PR   27-JAN-1994;    94US-0187166.
XX
PA   (GXBI-) GX BIOSYSTEMS AS.
XX
PI   Hasty DL,  Klemm P,  Molin S,  Pallesen L,  Sokurenko EV;
XX
DR   WPI; 1995-275442/36.
DR   P-PSDB; AAR76764.
XX
PT   Receptor specific bacterial adhesins - useful for targetting active
PT   compounds and microbial cells to locations of receptors
XX
PS   Example 5; Page 69-74; 152pp; English.
XX
CC   The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC   coli clinical isolates.  FimH is located at the tip of the type 1
CC   fimbriae and also intercalated at intervals in the fimbrial organelle.
CC   Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC   structures containing terminally located alpha-D-mannoside residues.
CC   FimH contains 4 cysteine residues assumed to direct folding of the
CC   molecule into distinct functional domains.  For comparison FimA and
CC   the minor components FimF and FimG only have 2 cysteine residues.
CC   The localisation of the cysteine residues in FimH points to a tandem
```

```
CC   arrangement of two ancestral genes.  Similar amino acids can be
CC   found in similar positions in the two halves of the FimH protein.  The
CC   "midway" point is located roughly around residue 150 in the mature
CC   protein.  The two halves or domains of FimH have evolved differently
CC   with the N-terminal section becoming the domain harbouring the receptor
CC   binding site, whereas the C-terminal sector became the domain of the
CC   molecule required for integration into the fimbrial organelle.  These
CC   sequences may be used in the production of variant FimH adhesins which
CC   may be useful for targetting active compounds and microbial cells to
CC   locations comprising selected receptors to which the adhesins bind.
XX
SQ   Sequence 900 BP; 204 A; 211 C; 240 G; 245 T; 0 other;

Alignment Scores:
Pred. No.:                4.93e-87        Length:          900
Score:                    814.00          Matches:         154
Percent Similarity:       97.52%          Conservative:    3
Best Local Similarity:    95.65%          Mismatches:      4
Query Match:              95.99%          Indels:          0
DB:                       16              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93075 (1-900)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::||||||||||||||||||:::||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACTACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db     559 TGTGATGTCACCGTTACTTTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621

RESULT 43
AAD29353
ID   AAD29353 standard; DNA; 837 BP.
XX
AC   AAD29353;
XX
DT   07-MAY-2002  (first entry)
XX
DE   Escherichia coli strain B212 FimH DNA.
XX
KW   FimH; immune response; antibacterial; enterobacillus-related disease;
KW   therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS   Escherichia coli B212.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..837
FT                   /*tag= a
FT                   /product= "E. coli FimH protein"
FT                   /transl_except= (pos:526..528, aa:Arg)
FT                   /transl_except= (pos:601..603, aa:His)
FT                   /note= "CDS does not include start and stop codon"
FT                   /partial
XX
PN   WO200204496-A2.
XX
PD   17-JAN-2002.
XX
PF   06-JUL-2001; 2001WO-US21525.
XX
PR   07-JUL-2000; 2000US-216750P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Langermann S,  Revel A,  Auguste C,  Burlein J;
XX
DR   WPI; 2002-171702/22.
DR   P-PSDB; AAE18414.
XX
PT   New immunogenic polypeptide, useful as vaccine for protecting against
PT   an enterobacillus-related disease in a patient at risk of contracting
PT   such disease, e.g. urinary tract infection or a bladder infection -
XX
PS   Example 1; Fig 1; 101pp; English.
XX
CC   The invention relates to bacterial immunogenic agents for administration
CC   to humans and non-human animals to stimulate an immune response. The
CC   invention also relates to methods for vaccination of mammalian species
CC   with variants of E. coli FimH protein derived from different strains of
CC   E. coli. The vaccine composition or the antibody is useful for protecting
CC   against and treating an enterobacillus-related disease in a patient
CC   afflicted or at a risk of contracting the disease. In particular, the
CC   disease is a urinary tract or bladder infection. The disease is caused
CC   by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC   The present sequence is Escherichia coli strain B212 FimH DNA.
XX
SQ   Sequence 837 BP; 191 A; 202 C; 224 G; 220 T; 0 other;

Alignment Scores:
Pred. No.:                7.7e-87         Length:          837
Score:                    812.00          Matches:         153
Percent Similarity:       96.89%          Conservative:    3
Best Local Similarity:    95.03%          Mismatches:      5
Query Match:              95.75%          Indels:          0
DB:                       24              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29353 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||   |||||||||:::||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGATCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCAGGCTCATTA 375
```

```
Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGCTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 44
AAD29370
ID AAD29370 standard; DNA; 837 BP.
XX
AC AAD29370;
XX
DT 07-MAY-2002 (first entry)
XX
DE Escherichia coli strain EC95 FimH DNA.
XX
KW FimH; immune response; antibacterial; enterobacillus-related disease;
KW therapy; vaccine; urinary tract infection; bladder; ds.
XX
OS Escherichia coli EC95.
XX

FH	Key	Location/Qualifiers
FT	CDS	1..837
FT		/*tag= a
FT		/product= "E. coli FimH protein"
FT		/transl_except= (pos:526..528, aa:Arg)
FT		/transl_except= (pos:601..603, aa:His)
FT		/note= "CDS does not include start and stop codon"
FT		/partial

XX
PN WO200204496-A2.
XX
PD 17-JAN-2002.
XX
PF 06-JUL-2001; 2001WO-US21525.
XX
PR 07-JUL-2000; 2000US-216750P.
XX
PA (MEDI-) MEDIMMUNE INC.
XX
PI Langermann S, Revel A, Auguste C, Burlein J;
XX
DR WPI; 2002-171702/22.
DR P-PSDB; AAE18431.
XX
PT New immunogenic polypeptide, useful as vaccine for protecting against
PT an enterobacillus-related disease in a patient at risk of contracting
PT such disease, e.g. urinary tract infection or a bladder infection -
XX
PS Example 1; Fig 1; 101pp; English.
XX
CC The invention relates to bacterial immunogenic agents for administration
CC to humans and non-human animals to stimulate an immune response. The
CC invention also relates to methods for vaccination of mammalian species
CC with variants of E. coli FimH protein derived from different strains of
CC E. coli. The vaccine composition or the antibody is useful for protecting
CC against and treating an enterobacillus-related disease in a patient
CC afflicted or at a risk of contracting the disease. In particular, the
CC disease is a urinary tract or bladder infection. The disease is caused
CC by a bacterium of the family Enterobacteriaceae, particularly E. coli.
CC The present sequence is Escherichia coli strain EC95 FimH DNA.
XX

SQ Sequence 837 BP; 194 A; 199 C; 223 G; 221 T; 0 other;

Alignment Scores:			
Pred. No.:	7.7e-87	Length:	837
Score:	812.00	Matches:	154
Percent Similarity:	97.52%	Conservative:	3
Best Local Similarity:	95.65%	Mismatches:	4
Query Match:	95.75%	Indels:	0
DB:	24	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AAD29370 (1-837)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||:::|||||||||||||||||||:::|||||||||||||||||||||||
Db    196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACTACCAGCGAAACGCCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||   ||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCTGGTGAGCAGTGCGGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 45
AAQ93069
ID AAQ93069 standard,
XX
AC AAQ93069;
XX
DT 15-MAR-1996 (firs
XX
DE FimH protein gene
XX
KW FimH; type 1 fimbr
KW FimA; FimF; FimG;
XX
OS Escherichia coli clinical isolate CI#10.
XX

FH	Key	Location/Qualifiers
FT	sig_peptide	1..63
FT		/*tag= a
FT	mat_peptide	64..888
FT		/*tag= b

XX
PN WO9520657-A1.
XX
PD 03-AUG-1995.
XX

0.

annoside residue;

PF 27-JAN-1995; 95WO-DK00042.
XX
PR 27-JAN-1994; 94US-0187166.
XX
PA (GXBI-) GX BIOSYSTEMS AS.
XX
PI Hasty DL, Klemm P, Molin S, Pallesen L, Sokurenko EV;
XX
DR WPI; 1995-275442/36.
DR P-PSDB; AAR76767.
XX
PT Receptor specific bacterial adhesins - useful for targetting active
PT compounds and microbial cells to locations of receptors
XX
PS Example 5; Page 69-74; 152pp; English.
XX
CC The sequences given in AAQ93062-75 encode FimH proteins from various E.
CC coli clinical isolates. FimH is located at the tip of the type 1
CC fimbriae and also intercalated at intervals in the fimbrial organelle.
CC Most forms of the FimH adhesin target to, and bind to, oligosaccharide
CC structures containing terminally located alpha-D-mannoside residues.
CC FimH contains 4 cysteine residues assumed to direct folding of the
CC molecule into distinct functional domains. For comparison FimA and
CC the minor components FimF and FimG only have 2 cysteine residues.
CC The localisation of the cysteine residues in FimH points to a tandem
CC arrangement of two ancestral genes. Similar amino acids can be
CC found in similar positions in the two halves of the FimH protein. The
CC "midway" point is located roughly around residue 150 in the mature
CC protein. The two halves or domains of FimH have evolved differently
CC with the N-terminal section becoming the domain harbouring the receptor
CC binding site, whereas the C-terminal sector became the domain of the
CC molecule required for integration into the fimbrial organelle. These
CC sequences may be used in the production of variant FimH adhesins which
CC may be useful for targetting active compounds and microbial cells to
CC locations comprising selected receptors to which the adhesins bind.
XX
SQ Sequence 888 BP; 203 A; 211 C; 234 G; 240 T; 0 other;

Alignment Scores:

Pred. No.:	3.29e-86	Length:	888
Score:	807.00	Matches:	155
Percent Similarity:	96.27%	Conservative:	0
Best Local Similarity:	96.27%	Mismatches:	2
Query Match:	95.17%	Indels:	4
DB:	16	Gaps:	1

US-09-900-575-29_COPY_26_186 (1-161) x AAQ93069 (1-888)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    139 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          |||||||||||||||||||||||||||||||||           ||||||||||||||||
Db    379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGT-----------AAAGCTGGCTCATTA 426

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 486

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    487 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 546

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db    547 CGGGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 606

Qy    161 Tyr 161
          |||
Db    607 TAT 609
```

Search completed: November 28, 2002, 19:14:35
Job time : 267 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2002, 19:06:46 ; Search time 2576 Seconds
(without alignments)
1818.925 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US09900575/runat_22112002_130708_4545/app_query.fasta_1.327
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09900575_@CGN_1_1_1616_@runat_22112002_130708_4545 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	848	100.0	837	6	AX354041	AX354041 Sequence
2	841	99.2	837	6	AX354051	AX354051 Sequence
3	841	99.2	2050	1	ECFIMFGH	X05672 E.coli gene
4	834	98.3	837	6	AX354040	AX354040 Sequence
5	834	98.3	837	6	AX354055	AX354055 Sequence
6	834	98.3	840	6	AX354044	AX354044 Sequence
7	834	98.3	840	6	AX354046	AX354046 Sequence
8	834	98.3	903	6	AX354082	AX354082 Sequence
9	834	98.3	903	6	AX363712	AX363712 Sequence
10	834	98.3	1040	1	AF154925	AF154925 Escherich
11	834	98.3	1040	1	AF154926	AF154926 Escherich
12	834	98.3	4637	6	AX357443	AX357443 Sequence
13	834	98.3	7416	6	AX354080	AX354080 Sequence
14	834	98.3	9299	6	AX357444	AX357444 Sequence
15	834	98.3	11313	1	AE000502	AE000502 Escherich
16	834	98.3	338534	1	ECOUW93	U14003 Escherichia
17	831	98.0	837	6	AX354088	AX354088 Sequence
18	830	97.9	837	6	AX354037	AX354037 Sequence
19	828	97.6	1040	1	AF154928	AF154928 Escherich
20	827	97.5	1040	1	AF306535	AF306535 Escherich
21	826	97.4	837	6	AX354049	AX354049 Sequence
22	826	97.4	1040	1	AF306536	AF306536 Escherich
23	825	97.3	840	6	AX354039	AX354039 Sequence
24	824	97.2	837	6	AX354054	AX354054 Sequence
25	824	97.2	14466	1	AE005662	AE005662 Escherich
26	824	97.2	168764	1	AP002569	AP002569 Escherich
27	823	97.1	1040	1	AF154927	AF154927 Escherich
28	822	96.9	837	6	AX354050	AX354050 Sequence
29	822	96.9	837	6	AX354052	AX354052 Sequence
30	822	96.9	937	1	AF288194	AF288194 Escherich
31	820	96.7	837	6	AX354042	AX354042 Sequence
32	819	96.6	837	6	AX354045	AX354045 Sequence
33	819	96.6	837	6	AX354056	AX354056 Sequence
34	819	96.6	840	6	AX354038	AX354038 Sequence
35	819	96.6	982	1	AF089840	AF089840 Escherich
36	819	96.6	1008	1	AF317710	AF317710 Escherich
37	819	96.6	1040	1	AF306537	AF306537 Escherich
38	818	96.5	837	6	AX354043	AX354043 Sequence
39	817	96.3	837	6	AX354035	AX354035 Sequence
40	816	96.2	837	6	AX354047	AX354047 Sequence
41	816	96.2	837	6	AX354048	AX354048 Sequence
42	816	96.2	3356	1	ECFIM	AJ225176 Escherich
43	812	95.8	837	6	AX354036	AX354036 Sequence
44	812	95.8	837	6	AX354053	AX354053 Sequence
45	807	95.2	1039	1	AF154929	AF154929 Escherich

ALIGNMENTS

RESULT 1

```
AX354041
LOCUS       AX354041                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 7 from Patent WO0204496.
ACCESSION   AX354041
VERSION     AX354041.1  GI:18618978
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 7 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      193 a    200 c    220 g    224 t
ORIGIN
```

Alignment Scores:

Pred. No.:	3.4e-78	Length:	837
Score:	848.00	Matches:	161
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AX354041 (1-837)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGTTGGTGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCGTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

```
RESULT 2
AX354051
LOCUS       AX354051                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 17 from Patent WO0204496.
ACCESSION   AX354051
VERSION     AX354051.1  GI:18618988
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 17 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      191 a    200 c    223 g    223 t
ORIGIN
```

Alignment Scores:

Pred. No.:	1.79e-77	Length:	837
Score:	841.00	Matches:	159
Percent Similarity:	99.38%	Conservative:	1
Best Local Similarity:	98.76%	Mismatches:	1
Query Match:	99.17%	Indels:	0
DB:	6	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AX354051 (1-837)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCGTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          . |||
Db    556 TAT 558
```

```
RESULT 3
ECFIMFGH
LOCUS       ECFIMFGH                2050 bp    DNA     linear   BCT 12-SEP-1993
DEFINITION  E.coli genes fimF, fimG and fimH.
ACCESSION   X05672
VERSION     X05672.1  GI:41463
KEYWORDS    fimF gene; fimG gene; fimH gene; regulatory protein.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
```

```
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 2050)
  AUTHORS   Klemm,P. and Christiansen,G.
  TITLE     Three fim genes required for the regulation of length and mediation
            of adhesion of Escherichia coli type 1 fimbriae
  JOURNAL   Mol. Gen. Genet. 208 (3), 439-445 (1987)
  MEDLINE   88038337
   PUBMED   2890081
FEATURES             Location/Qualifiers
     source          1. .2050
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     RBS             17. .21
                     /note="ribosome binding site (fimF)"
     CDS             28. .558
                     /note="pot.fimF precursor (AA 1-176)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA29154.1"
                     /db_xref="GI:41464"
                     /db_xref="SWISS-PROT:P08189"
                     /translation="MRNKPFYLLCAFLWLAVSHALAADSTITIRGYVRDNGCSVAAES
                     TNFTVDLMENAAKQFNNIGATTPVVSFRILLLPCGNVVSRRKGWVYWRADSHNANLLA
                     LENTASAASGLGIQLLNEQQNQIPLNAPSSALSWTTLTPGKPNTLNFYARLMATQVPV
                     TAGHINATATFTLEYQ"
     misc_feature    87. .88
                     /note="pot.processing site of the mature protein"
     RBS             561. .565
                     /note="ribosome binding site (fimG)"
     CDS             571. .1074
                     /note="pot.fimG precursor (AA 1-167)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA29155.1"
                     /db_xref="GI:41465"
                     /db_xref="SWISS-PROT:P08190"
                     /translation="MKWCKRGYVLAAILALASATIQAADVTITVNGKVVAKPCTVSTT
                     NATVDLGDLYSFSLMSSGAASAWHDVALELTNCPVGTSRVTASFSGAADSTGYYKNQG
                     TAQNIQLELQDDSGNTLNTGATKTVQVDDSSQSAHFPLQVRALTVKVGATQGTIQAVI
                     SITYTYS"
     misc_feature    639. .640
                     /note="pot.processing site of the mature protein"
     RBS             1081. .1085
                     /note="ribosome binding site (fimH)"
     CDS             1094. .1996
                     /note="pot.fimH precursor (AA 1-300)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA29156.1"
                     /db_xref="GI:41466"
                     /db_xref="SWISS-PROT:P08191"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYRGSVPIPLTVYCAKSQNLGYYL
                     SGTHADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN
                     YARTGGQVTAGNVQSIIGVTFVYQ"
     misc_feature    1162. .1163
                     /note="pot.processing site of the mature protein"
BASE COUNT      494 a    489 c    542 g    525 t
ORIGIN

Alignment Scores:
Pred. No.:                5.01e-77        Length:          2050
Score:                    841.00          Matches:         159
Percent Similarity:       99.38%          Conservative:    1
Best Local Similarity:    98.76%          Mismatches:      1
Query Match:              99.17%          Indels:          0
DB:                       1               Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x ECFIMFGH (1-2050)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 1291

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 1351

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 1411

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 1471

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db    1472 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 1531

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1532 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 1591

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1592 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 1651

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1652 CGTGATGTCACCGTTACTCTGCCGGACTACCGTGGTTCAGTGCCAATTCCTCTTACCGTT 1711

Qy     161 Tyr 161
           |||
Db    1712 TAT 1714

RESULT 4
AX354040
LOCUS       AX354040                 837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 6 from Patent WO0204496.
ACCESSION   AX354040
VERSION     AX354040.1  GI:18618977
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 6 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      192 a    202 c    223 g    220 t
ORIGIN

Alignment Scores:
Pred. No.:                9.4e-77         Length:          837
Score:                    834.00          Matches:         158
Percent Similarity:       98.76%          Conservative:    1
Best Local Similarity:    98.14%          Mismatches:      2
Query Match:              98.35%          Indels:          0
DB:                       6               Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354040 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135
```

```
Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGATTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGTGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGTGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

```
RESULT 5
AX354055
LOCUS       AX354055                 837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 21 from Patent WO0204496.
ACCESSION   AX354055
VERSION     AX354055.1  GI:18618992
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 21 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      191 a    201 c    222 g    223 t
ORIGIN

Alignment Scores:
Pred. No.:              9.4e-77       Length:         837
Score:                  834.00        Matches:        158
Percent Similarity:     98.76%        Conservative:   1
Best Local Similarity:  98.14%        Mismatches:     2
Query Match:            98.35%        Indels:         0
DB:                     6             Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AX354055 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195
```

```
Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

```
RESULT 6
AX354044
LOCUS       AX354044                 840 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 10 from Patent WO0204496.
ACCESSION   AX354044
VERSION     AX354044.1  GI:18618981
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 10 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      193 a    202 c    224 g    221 t
ORIGIN

Alignment Scores:
Pred. No.:              9.44e-77      Length:         840
Score:                  834.00        Matches:        158
Percent Similarity:     98.76%        Conservative:   1
Best Local Similarity:  98.14%        Mismatches:     2
Query Match:            98.35%        Indels:         0
DB:                     6             Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AX354044 (1-840)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGATTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255
```

```
Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGTGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGTGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

```
RESULT 7
AX354046
LOCUS       AX354046                840 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 12 from Patent WO0204496.
ACCESSION   AX354046
VERSION     AX354046.1  GI:18618983
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 12 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      193 a    201 c    222 g    224 t
ORIGIN

Alignment Scores:
Pred. No.:                9.44e-77      Length:          840
Score:                    834.00        Matches:         158
Percent Similarity:       98.76%        Conservative:    1
Best Local Similarity:    98.14%        Mismatches:      2
Query Match:              98.35%        Indels:          0
DB:                       6             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354046 (1-840)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

```
RESULT 8
AX354082
LOCUS       AX354082                903 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 48 from Patent WO0204496.
ACCESSION   AX354082
VERSION     AX354082.1  GI:18618996
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 48 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .903
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Sequence J96 fimH plus native signal sequence"
BASE COUNT      206 a    213 c    240 g    244 t
ORIGIN

Alignment Scores:
Pred. No.:                1.03e-76      Length:          903
Score:                    834.00        Matches:         158
Percent Similarity:       98.76%        Conservative:    1
Best Local Similarity:    98.14%        Mismatches:      2
Query Match:              98.35%        Indels:          0
DB:                       6             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354082 (1-903)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438
```

```
Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db    559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy    161 Tyr 161
          |||
Db    619 TAT 621
```

RESULT 9
AX363712
LOCUS AX363712 903 bp DNA linear PAT 15-FEB-2002
DEFINITION Sequence 207 from Patent EP1178052.
ACCESSION AX363712
VERSION AX363712.1 GI:18695826
KEYWORDS .
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

FEATURES	Location/Qualifiers
source	1. .903
	/organism="Escherichia coli"
	/db_xref="taxon:562"

BASE COUNT 206 a 213 c 240 g 244 t
ORIGIN

Alignment Scores:

Pred. No.:	1.03e-76	Length:	903
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	6	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AX363712 (1-903)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    139 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||||
Db    379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
```

```
          ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db    559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy    161 Tyr 161
          |||
Db    619 TAT 621
```

RESULT 10
AF154925
LOCUS AF154925 1040 bp DNA linear BCT 10-JUL-2000
DEFINITION Escherichia coli FimH precursor (fimH) gene, fimH-241 allele, complete cds.
ACCESSION AF154925
VERSION AF154925.1 GI:5524627
KEYWORDS .
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1040)
AUTHORS Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R., Russell,P.W. and Orndorff,P.E.
TITLE Genetic characterization of Escherichia coli type 1 pilus adhesin mutants and identification of a novel binding phenotype
JOURNAL J. Bacteriol. 182 (14), 4012-4021 (2000)
MEDLINE 20327582
PUBMED 10869080
REFERENCE 2 (bases 1 to 1040)
AUTHORS Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R., Russell,P.W. and Orndorff,P.E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1999) Microbiology, Pathology, and Parasitology, North Carolina State University, College of Veterinary Medicine, 4700 Hillsborough Street, Raleigh, NC 27606, USA

FEATURES	Location/Qualifiers
source	1. .1040
	/organism="Escherichia coli"
	/strain="K12"
	/db_xref="taxon:562"
gene	47. .949
	/gene="fimH"
	/note="mutant allele"
	/allele="fimH-241"
CDS	47. .949
	/gene="fimH"
	/note="type 1 pili adhesin"
	/codon_start=1
	/transl_table=11
	/product="FimH precursor"
	/protein_id="AAD44319.1"
	/db_xref="GI:5524628"
	/translation="MKRVITLFAVLLMGWSVNDWSFACKTANGTAIPIGGGSANVYVN LAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSS YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN YARTGGQVTAGNVQSIIGVTFVYQ"
variation	102
	/gene="fimH"
	/note="confers an alanine to aspartic acid change near the signal sequence processing site"
	/replace="c"

BASE COUNT 253 a 243 c 270 g 274 t
ORIGIN

Alignment Scores:

Pred. No.:	1.21e-76	Length:	1040
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	1	Gaps:	0

```
US-09-900-575-29_COPY_26_186 (1-161) x AF154925 (1-1040)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 244

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 304

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 364

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 424

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db     425 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 484

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     485 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 544

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 604

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db     605 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 664

Qy     161 Tyr 161
           |||
Db     665 TAT 667

RESULT 11
AF154926
LOCUS       AF154926               1040 bp    DNA     linear   BCT 10-JUL-2000
DEFINITION  Escherichia coli FimH precursor (fimH) gene, fimH-236 allele,
            complete cds.
ACCESSION   AF154926
VERSION     AF154926.1  GI:5524629
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1040)
  AUTHORS   Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R.,
            Russell,P.W. and Orndorff,P.E.
  TITLE     Genetic characterization of Escherichia coli type 1 pilus adhesin
            mutants and identification of a novel binding phenotype
  JOURNAL   J. Bacteriol. 182 (14), 4012-4021 (2000)
  MEDLINE   20327582
   PUBMED   10869080
REFERENCE   2  (bases 1 to 1040)
  AUTHORS   Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R.,
            Russell,P.W. and Orndorff,P.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Microbiology, Pathology, and Parasitology,
            North Carolina State University, College of Veterinary Medicine,
            4700 Hillsborough Street, Raleigh, NC 27606, USA
FEATURES             Location/Qualifiers
     source          1. .1040
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
     gene            47. .949
                     /gene="fimH"
                     /note="mutant allele"
                     /allele="fimH-236"
     CDS             47. .949
                     /gene="fimH"
                     /note="type 1 pili adhesin"
                     /codon_start=1
                     /transl_table=11
                     /product="FimH precursor"
                     /protein_id="AAD44320.1"
                     /db_xref="GI:5524630"
                     /translation="MKRVITLFAVLLMGWSVNAWLFACKTANGTAIPIGGGSANVYVN
                     LAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL
                     SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN
                     YARTGGQVTAGNVQSIIGVTFVYQ"
     variation       108
                     /gene="fimH"
                     /note="confers a serine to leucine change at the signal
                     sequence processing site"
                     /replace="c"
BASE COUNT      252 a    243 c    270 g    275 t
ORIGIN

Alignment Scores:
Pred. No.:                1.21e-76        Length:              1040
Score:                    834.00          Matches:             158
Percent Similarity:       98.76%          Conservative:        1
Best Local Similarity:    98.14%          Mismatches:          2
Query Match:              98.35%          Indels:              0
DB:                       1               Gaps:                0

US-09-900-575-29_COPY_26_186 (1-161) x AF154926 (1-1040)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 244

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 304

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 364

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 424

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db     425 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 484

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     485 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 544

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 604

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db     605 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 664

Qy     161 Tyr 161
           |||
Db     665 TAT 667

RESULT 12
AX357443
LOCUS       AX357443               4637 bp    DNA     linear   PAT 13-FEB-2002
DEFINITION  Sequence 167 from Patent WO0185208.
```

```
ACCESSION   AX357443
VERSION     AX357443.1  GI:18674526
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Sebbel,P., Dunant,N., Bachmann,M., Tissot,A. and Lechener,F.
  TITLE     Molecular antigen arrays and vaccines
  JOURNAL   Patent: WO 0185208-A 167 15-NOV-2001;
            Cytos Biotechnology AG (CH) ; Sebbel, Peter (CH) ; Dunant, Nicolas
            (CH) ; Bachmann, Martin (CH) ; Tissot, Alain (CH) ; Lechener,
            Franziska (CH)
FEATURES             Location/Qualifiers
     source          1. .4637
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="pFIMFGH"
BASE COUNT     1124 a   1156 c   1188 g   1169 t
ORIGIN

Alignment Scores:
Pred. No.:                 6.74e-76      Length:          4637
Score:                     834.00        Matches:         158
Percent Similarity:        98.76%        Conservative:    1
Best Local Similarity:     98.14%        Mismatches:      2
Query Match:               98.35%        Indels:          0
DB:                        6             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX357443 (1-4637)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3441 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 3500

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3501 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 3560

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3561 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 3620

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3621 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 3680

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||||
Db    3681 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 3740

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3741 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 3800

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3801 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 3860

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    3861 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 3920

Qy     161 Tyr 161
           |||
Db    3921 TAT 3923

RESULT 13
AX354080
LOCUS       AX354080               7416 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 46 from Patent WO0204496.
ACCESSION   AX354080
VERSION     AX354080.1  GI:18618994
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 46 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .7416
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="Sequence of Plasmid pCGA126-1"
BASE COUNT     1858 a   1807 c   1908 g   1843 t
ORIGIN

Alignment Scores:
Pred. No.:                 1.16e-75      Length:          7416
Score:                     834.00        Matches:         158
Percent Similarity:        98.76%        Conservative:    1
Best Local Similarity:     98.14%        Mismatches:      2
Query Match:               98.35%        Indels:          0
DB:                        6             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354080 (1-7416)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1244 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 1303

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1304 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 1363

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 1423

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1424 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 1483

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||||
Db    1484 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 1543

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1544 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 1603

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1604 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 1663

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    1664 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 1723

Qy     161 Tyr 161
           |||
Db    1724 TAT 1726

RESULT 14
AX357444
LOCUS       AX357444               9299 bp    DNA     linear   PAT 13-FEB-2002
DEFINITION  Sequence 168 from Patent WO0185208.
ACCESSION   AX357444
VERSION     AX357444.1  GI:18674527
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequences.
```

```
REFERENCE   1
  AUTHORS   Sebbel,P., Dunant,N., Bachmann,M., Tissot,A. and Lechener,F.
  TITLE     Molecular antigen arrays and vaccines
  JOURNAL   Patent: WO 0185208-A 168 15-NOV-2001;
            Cytos Biotechnology AG (CH) ; Sebbel, Peter (CH) ; Dunant, Nicolas
            (CH) ; Bachmann, Martin (CH) ; Tissot, Alain (CH) ; Lechener,
            Franziska (CH)
FEATURES             Location/Qualifiers
     source          1. .9299
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="pFIMAICDFGH"
BASE COUNT     2371 a    2185 c    2410 g    2333 t
ORIGIN

Alignment Scores:
Pred. No.:                 1.5e-75       Length:          9299
Score:                     834.00        Matches:         158
Percent Similarity:        98.76%        Conservative:    1
Best Local Similarity:     98.14%        Mismatches:      2
Query Match:               98.35%        Indels:          0
DB:                        6             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX357444 (1-9299)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8109 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 8168

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8169 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 8228

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8229 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 8288

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8289 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 8348

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||
Db    8349 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 8408

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8409 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 8468

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8469 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 8528

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db    8529 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 8588

Qy     161 Tyr 161
           |||
Db    8589 TAT 8591

RESULT 15
AE000502
LOCUS       AE000502               11313 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 392 of 400 of the complete
            genome.
ACCESSION   AE000502 U00096
VERSION     AE000502.1  GI:2367374
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.

REFERENCE   1  (bases 1 to 11313)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11313)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11313)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11313)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     On Sep 9, 1997 this sequence version replaced gi:1790766.
            This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11313
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     protein_bind    429. .443
                     /note="central position to fimB promoter: -211.5"
                     /bound_moiety="Lrp documented site"
     promoter        613. .642
                     /note="factor Sigma70; predicted +1 start at 4538423"
     gene            751. .1353
                     /gene="fimB"
                     /note="b4312"
     CDS             751. .1353
                     /gene="fimB"
                     /function="regulator; Surface structures"
```

/note="o200; 100 pct identical to FIMB_ECOLI SW: P04742;
CG Site No. 18355; alternate gene name pil"
/codon_start=1
/transl_table=11
/product="recombinase involved in phase variation;
regulator for fimA"
/protein_id="AAC77268.1"
/db_xref="GI:1790767"
/translation="MKNKADNKKRNFLTHSEIESLLKAANTGPHAARNYCLTLLCFIH
GFRASEICRLRISDIDLKAKCIYIHRLKKGFSTTHPLLNKEVQALKNWLSIRTSYPHA
ESEWVFLSRKGNPLSRQQFYHIISTSGGNAGLSLEIHPHMLRHSCGFALANMGIDTRL
IQDYLGHRNIRHTVWYTASNAGRFYGIWDRARGRQRHAVL"

gene 1831. .2427
/gene="fimE"
/note="b4313"

CDS 1831. .2427
/gene="fimE"
/function="regulator; Surface structures"
/note="o198; 100 pct identical to FIME_ECOLI SW: P04741;
CG Site No. 18346"
/codon_start=1
/transl_table=11
/product="recombinase involved in phase variation;
regulator for fimA"
/protein_id="AAC77269.1"
/db_xref="GI:1790768"
/translation="MSKRRYLTGKEVQAMMQAVCYGATGARDYCLILLAYRHGMRISE
LLDLHYQDLDLNEGRINIRRLKNGFSTVHPLRFDEREAVERWTQERANWKGADRTDAI
FISRRGSRLSRQQAYRIIRDAGIEAGTVTQTHPHMLRHACGYELAERGADTRLIQDYL
GHRNIRHTVRYTASNAARFAGLWERNNLINEKLKREEV"

gene 2909. .3457
/gene="fimA"
/note="b4314"

CDS 2909. .3457
/gene="fimA"
/function="structural component; Surface structures"
/note="o182; 100 pct identical to FM1A_ECOLI SW: P04128;
CG Site No. 18358; alternate gene names fimD, pilA"
/codon_start=1
/transl_table=11
/product="major type 1 subunit fimbrin (pilin)"
/protein_id="AAC77270.1"
/db_xref="GI:1790769"
/translation="MKIKTLAIVVLSALSLSSTAALAAATTVNGGTVHFKGEVVNAAC
AVDAGSVDQTVQLGQVRTASLAQEGATSSAVGFNIQLNDCDTNVASKAAVAFLGTAID
AGHTNVLALQSSAAGSATNVGVQILDRTGAALTLDGATFSSETTLNNGTNTIPFQARY
FATGAATPGAANADATFKVQYQ"

gene 3414. .4061
/gene="fimI"
/note="b4315"

CDS 3414. .4061
/gene="fimI"
/function="structural component; Surface structures"
/note="o215"
/codon_start=1
/transl_table=11
/product="fimbrial protein"
/protein_id="AAC77271.1"
/db_xref="GI:1790770"
/translation="MLLMRMRPSRFSINNLPRFRDVITGRDAHPCAIKITMKRKRLFL
LASLLPMFALAGNKWNTTLPGGNMQFQGVIIAETCRIEAGDKQMTVNMGQISSNRFHA
VGEDSAPVPFVIHLRECSTVVSERVGVAFHGVADGKNPDVLSVGEGPGTATNIGVALF
DDEGNLVPINRPPANWKRLYSGSTSLHFIAKYRATGRRVTGGIANAQAWFSLTYQ"

gene 4098. .4823
/gene="fimC"
/note="b4316"

CDS 4098. .4823
/gene="fimC"
/function="factor; Surface structures"
/note="o241; 100 pct identical to FIMC_ECOLI SW: P31697;
CG Site No. 18352; alternate gene name pil"
/codon_start=1
/transl_table=11
/product="periplasmic chaperone, required for type 1
fimbriae"
/protein_id="AAC77272.1"
/db_xref="GI:1790771"
/translation="MSNKNVNVRKSQEITFCLLAGILMFMAMMVAGRAEAGVALGATR
VIYPAGQKQEQLAVTNNDENSTYLIQSWVENADGVKDGRFIVTPPLFAMKGKKENTLR
ILDATNNQLPQDRESLFWMNVKAIPSMDKSKLTENTLQLAIISRIKLYYRPAKLALPP
DQAAEKLRFRRSANSLTLINPTPYYLTVTELNAGTRVLENALVPPMGESTVKLPSDAG
SNITYRTINDYGALTPKMTGVME"

gene 4891. .7527
/gene="fimD"
/note="b4317"

CDS 4891. .7527
/gene="fimD"
/function="membrane; Outer membrane constituents"
/note="o878; 100 pct identical to FIMD_ECOLI SW: P30130;
CG Site No. 18349"
/codon_start=1
/transl_table=11
/product="outer membrane protein; export and assembly of
type 1 fimbriae, interrupted"
/protein_id="AAC77273.1"
/db_xref="GI:1790772"
/translation="MSYLNLRLYQRNTQCLHIRKHRLAGFFVRLVVACAFAAQAPLSS
ADLYFNPRFLADDPQAVADLSRFENGQELPPGTYRVDIYLNNGYMATRDVTFNTGDSE
QGIVPCLTRAQLASMGLNTASVAGMNLLADDACVPLTTMVQDATAHLDVGQQRLNLTI
PQAFMSNRARGYIPPELWDPGINAGLLNYNFSGNSVQNRIGGNSHYAYLNLQSGLNIG
AWRLRDNTTWSYNSSDRSSGSKNKWQHINTWLERDIIPLRSRLTLGDGYTQGDIFDGI
NFRGAQLASDDNMLPDSQRGFAPVIHGIARGTAQVTIKQNGYDIYNSTVPPGPFTIND
IYAAGNSGDLQVTIKEADGSTQIFTVPYSSVPLLQREGHTRYSITAGEYRSGNAQQEK
TRFFQSTLLHGLPAGWTIYGGTQLADRYRAFNFGIGKNMGALGALSVDMTQANSTLPD
DSQHDGQSVRFLYNKSLNESGTNIQLVGYRYSTSGYFNFADTTYSRMNGYNIETQDGV
IQVKPKFTDYYNLAYNKRGKLQLTVTQQLGRTSTLYLSGSHQTYWGTSNVDEQFQAGL
NTAFEDINWTLSYSLTKNAWQKGRDQMLALNVNIPFSHWLRSDSKSQWRHASASYSMS
HDLNGRMTNLAGVYGTLLEDNNLSYSVQTGYAGGGDGNSGSTGYATLNYRGGYGNANI
GYSHSDDIKQLYYGVSGGVLAHANGVTLGQPLNDTVVLVKAPGAKDAKVENQTGVRTD
WRGYAVLPYATEYRENRVALDTNTLADNVDLDNAVANVVPTRGAIVRAEFKARVGIKL
LMTLTHNNKPLPFGAMVTSESSQSSGIVADNGQVYLSGMPLAGKVQVKWGEEENAHCV
ANYQLPPESQQQLLTQLSAECR"

gene 7537. .8067
/gene="fimF"
/note="b4318"

CDS 7537. .8067
/gene="fimF"

Alignment Scores:

Pred. No.:	1.88e-75	Length:	11313
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	1	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AE000502 (1-11313)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8741 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 8800

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8801 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 8860

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8861 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 8920

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8921 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 8980

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
```

```
Db    8981 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 9040

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9041 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 9100

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9101 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 9160

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db    9161 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 9220

Qy     161 Tyr 161
           |||
Db    9221 TAT 9223
```

RESULT 16
ECOUW93

LOCUS	ECOUW93 338534 bp DNA linear BCT 30-JAN-2001
DEFINITION	Escherichia coli K-12 chromosomal region from 92.8 to 00.1 minutes.
ACCESSION	U14003
VERSION	U14003.1 GI:1263172
KEYWORDS	.
SOURCE	Escherichia coli.
ORGANISM	Escherichia coli Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE	1 (bases 1 to 338534)
AUTHORS	Burland,V., Plunkett,G. III, Sofia,H.J., Daniels,D.L. and Blattner,F.R.
TITLE	Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes
JOURNAL	Nucleic Acids Res. 23 (12), 2105-2119 (1995)
MEDLINE	95334362
PUBMED	7610040
REFERENCE	2 (bases 1 to 338534)
AUTHORS	Plunkett,G.I.I.I.
TITLE	Direct Submission
JOURNAL	Submitted (22-AUG-1994) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
COMMENT	On Apr 17, 1996 this sequence version replaced gi:536929. This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. The entire sequence was independently determined from E. coli MG1655; overlaps with other sequence determinations are annotated. The start of this entry overlaps the end of the entry ECOUW89 (U00006) by 1885 bp.

FEATURES Location/Qualifiers

source	1. .338534 /organism="Escherichia coli" /strain="K-12" /sub_strain="MG1655" /db_xref="taxon:562" /map="92.8 to 100 minutes" /note="This sequence comprises the following lambda clones: DD928(EC27-278), DD930(EC30-188), DD933(EC27-1320), DD935(EC30MM32), DD937(EC27-1070), DD941(EC17-142), DD945(EC17-8), DD947(EC24A-34), DD949(EC22-169), DD952(EC27-297), DD953(EC27-409), DD956(EC17-103), DD958(EC30MM1), DD960(EC30MM2), DD962(EC21-104), DD965(EC23A-40), DD968(EC30K660A-4pp), DD970(EC19-202), DD974(EC19-61), DD975(EC18-233), DD977(EC30MM7), DD980(EC27-1151), DD981(EC18-3), DD984(EC17-101), DD987(EC18-115), DD990(EC17-136), DD992(EC18-282), DD995(EC17-115), DD997(EC18-113); M13mp19 or Janus vectors were used for subcloning"
misc_feature	1. .1885 /note="1885 bp overlap with end of GenBank Accession Number U00006 (ECOWU89)"
CDS	complement(<1. .398) /note="ORF_f326 of GenBank Accession Number U00006 (ECOWU89)" /codon_start=1 /transl_table=11 /label=ORF_f132p /protein_id="AAA96985.1" /db_xref="GI:1263173" /translation="MGFTTRVKSEASEKKPFNFALFWDKYGTFFILAIIVAIFGSLSP EYFLTTNNITQIFVQSSVTVLIGMGEFFAILVAGIDLSVGAILALSGMVTAKLMLAGV DPFLAAMIGGVLVGGALGAINGCLVNWTGL"
misc_structure	277. .397 /note="predicted bend of 75.10 degrees"
CDS	complement(377. .1909) /note="ORF_f510" /codon_start=1 /transl_table=11 /label=ORF_f510 /protein_id="AAA96986.1" /db_xref="GI:536931" /translation="MATPYISMAGIGKSFGPVHALKSVNLTVYPGEIHALLGENGAGK STLMKVLSGIHEPTKGTITINNISYNKLDHKLAAQLGIGIIYQELSVIDELTVLENLY IGRHLTKKICGVNIIDWREMRVRAAMMLLRVGLKVDLDEKVANLSISHKQMLEIAKTL MLDAKVIIMDEPTSSLTNKEVDYLFLIMNQLRKEGTAIVYISHKLAEIRRICDRYTVM KDGSSVCSGIVSDVSNDDIVRLMVGRELQNRFNAMKENVSNLAHETVFEVRNVTSRDR KKVRDISFSVCRGEILGFAGLVGSGRTELMNCLFGVDKRAGGEIRLNGKDISPRSPLD AVKKGMAYITESRRDNGFFPNFSIAQNMAISRSLKDGGYKGAMGLFHEVDEQRTAENQ RELLALKCHSVNQNITELSGGNQQKVLISKWLCCCPEVIIFDEPTRGIDVGAKAEIYK VMRQLADDGKVILMVSSELPEIITVCDRIAVFCEGRLTQILTNRDDMSEEEIMAWALP QE"
misc_structure	809. .931 /note="predicted bend of 74.75 degrees"
misc_structure	1768. .1889 /note="predicted bend of 75.87 degrees"
promoter	complement(1983. .2010) /note="promoter matrix score of 55; putative; resembles flh-regulated promoter"
repeat_region	complement(1985. .2021) /standard_name="REP; repetitive extragenic palindromic element" /note="contains 1 REP sequence"
CDS	complement(2036. .2971) /note="ORF_f311" /codon_start=1 /transl_table=11 /label=ORF_f311 /protein_id="AAA96987.1" /db_xref="GI:536932" /translation="MNKYLKYFSGTLVGLMLSTSAFAAAEYAVVLKTLSNPFWVDMKK GIEDEAKTLGVSVDIFASPSEGDFQSQLQLFEDLSNKNYKGIAFAPLSSVNLVMPVAR AWKKGIYLVNLDEKIDMDNLKKAGGNVEAFVTTDNVAVGAKGASFIIDKLGAEGGEVA IIEGKAGNASGEARRNGATEAFKKASQIKLVASQPADWDRIKALDVATNVLQRNPNIK AIYCANDTMAMGVAQAVANAGKTGKVLVVGTDGIPEARKMVEAGQMTATVAQNPADIG ATGLKLMVDAEKSGKVIPLDKAPEFKLVDSILVTQ"
misc_structure	3010. .3121 /note="predicted bend of 73.22 degrees"
CDS	complement(3030. .3953) /note="ORF_f307" /codon_start=1 /transl_table=11 /label=ORF_f307 /protein_id="AAA96988.1" /db_xref="GI:536933" /translation="MASSIQGQENVMSQSEFDSALPNGIGLAPYLRMKQEGMTENESR IVEWLLKPGNLSCAPAIKDVAEALAVSEAMIVKVSKLLGFSGFRNLRSALEDYFSQSE QVLPSELAFDEAPQDVVNKVFNITLRTIMEGQSIVNVDEIHRAARFFYQARQRDLYGA GGSNAICADVQHKFLRIGVRCQAYPDAHIMMMSASLLQEGDVVLVVTHSGRTSDVKAA VELAKKNGAKIICITHSYHSPIAKLADYIICSPAPETPLLGRNASARILQLTLLDAFF VSVAQLNIEQANINMQKTGAIVDFFSPGALK"
promoter	complement(3952. .3979) /note="promoter matrix score of 61; putative"
misc_structure	4151. .4251

```
                /note="predicted bend of 72.06 degrees"
gene            4279. .4728
                /gene="yjcA"
CDS             4279. .4728
                /gene="yjcA"
                /codon_start=1
                /transl_table=11
                /label=ORF_o149
                /protein_id="AAA96989.1"
                /db_xref="GI:536934"
                /translation="MKKIAFGCDHVGFILKHEIVAHLVERGVEVIDKGTWSSERTDYP
                HYASQVALAVAGGEVDGGILICGTGVGISIAANKFAGIRAVVCSEPYSAQLSRQHNDT
                NVLAFGSRVVGLELAKMIVDAWLGAQYEGGRHQQRVEAITAIEQRRN"
misc_feature    complement(4520. .16191)
                /note="corresponds to GenBank Accession Number D90227
                (ECOPHN) 1. .11672; the related sequence GenBank Accession
                Number J05260 (ECOPHNAQ) is from E. coli B, not K-12"
gene            complement(4739. .5146)
                /gene="phnQ"
CDS             complement(4739. .5146)
                /gene="phnQ"
                /note="annotated in GenBank Accession Number J05260
                (ECOPHNAQ), but probably not a gene"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA96990.1"
                /db_xref="GI:536935"
                /translation="MPAPYPTSSPILIFHSPLFILQLPTLLIKLALAIVQLPAAIFKL
                AAAVVILSAPIVAHFTLPAVVARIFIVSPTVRVISVINRLLNIANGVVDNIGLGRNVD
                SVSRVNKRNSGKRFISQPQKETLSTLIITHPCK"
gene            complement(5272. .6030)
                /gene="phnP"
CDS             complement(5272. .6030)
                /gene="phnP"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA96991.1"
                /db_xref="GI:536936"
                /translation="MSLTLTLTGTGGAQGVPAWGCECAACARARRSPQYRRQPCSGVV
                KFNDAITLIDAGLHDLADRWSPGSFQQFLLTHYHMDHVQGLFPLRWGVGDPIPVYGPP
                DEQGCDDLFKHPGLLDFSHTVEPFVVFDLQGLQVTPLPLNHSKLTFGYLLETAHSRVA
                WLSDTAGLPEKTLKFLRNNQPQVMVMDCSHPPRADAPRNHCDLNTVLALNQVIRSPRV
                ILTHISHQFDAWLMENALPSGFEVGFDGMEIGVA"
misc_structure  5452. .5586
                /note="predicted bend of 75.60 degrees"
gene            complement(6032. .6466)
                /gene="phnO"
CDS             complement(6032. .6466)
                /gene="phnO"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA96992.1"
                /db_xref="GI:536937"
                /translation="MPACELRPATQYDTDAVYALICELKQAEFDHHAFRVGFNANLRD
                PNMRYHLALLDGEVVGMIGLHLQFHLHHVNWIGEIQELVVMPQARGLNVGSKLLAWAE
                EEARQAGAEMTELSTNVKRHDAHRFYLREGYEQSHFRFTKAL"
gene            complement(6453. .7010)
                /gene="phnN"
CDS             complement(6453. .7010)
                /gene="phnN"
                /note="HisP-like nucleotide binding protein (ATP-binding
                protein)"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA96993.1"
                /db_xref="GI:536938"
                /translation="MMGKLIWLMGPSGSGKDSLLAELRLREQTQLLVAHRYITRDASA
                GSENHIALSEQEFFTRAGQNLLALSWHANGLYYGVGVEIDLWLHAGFDVLVNGSRAHL
                PQARARYQSALLPVCLQVSPEILRQRLENRGRENASEINARLARAARYTPQDCHTLNN
                DGSLRQSVDTLLTLIHQKEKHHACL"
gene            complement(7010. .8146)
                /gene="phnM"
CDS             complement(7010. .8146)
                /gene="phnM"

Alignment Scores:
Pred. No.:              9.42e-74        Length:         338534
Score:                  834.00          Matches:        158
Percent Similarity:     98.76%          Conservative:   1
Best Local Similarity:  98.14%          Mismatches:     2
Query Match:            98.35%          Indels:         0
DB:                     1               Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x ECOUW93 (1-338534)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  239779 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 239838

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  239839 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 239898

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  239899 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 239958

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  239959 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 240018

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||
Db  240019 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 240078

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  240079 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 240138

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  240139 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 240198

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db  240199 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 240258

Qy     161 Tyr 161
           |||
Db  240259 TAT 240261

RESULT 17
AX354088
LOCUS       AX354088                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 54 from Patent WO0204496.
ACCESSION   AX354088
VERSION     AX354088.1  GI:18619002
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 54 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      192 a    201 c    221 g    223 t
ORIGIN

Alignment Scores:
Pred. No.:              1.91e-76        Length:         837
```

```
Score:                  831.00          Matches:        157
Percent Similarity:     98.76%          Conservative:   2
Best Local Similarity:  97.52%          Mismatches:     2
Query Match:            98.00%          Indels:         0
DB:                     6               Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AX354088 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 18
AX354037
LOCUS       AX354037                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 3 from Patent WO0204496.
ACCESSION   AX354037
VERSION     AX354037.1  GI:18618974
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 3 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      193 a    201 c    221 g    222 t
ORIGIN

Alignment Scores:
Pred. No.:              2.43e-76        Length:         837
Score:                  830.00          Matches:        157
Percent Similarity:     98.14%          Conservative:   1
Best Local Similarity:  97.52%          Mismatches:     3
Query Match:            97.88%          Indels:         0
DB:                     6               Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AX354037 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 19
AF154928
LOCUS       AF154928                1040 bp    DNA     linear   BCT 10-JUL-2000
DEFINITION  Escherichia coli FimH precursor (fimH) gene, fimH-205 allele,
            complete cds.
ACCESSION   AF154928
VERSION     AF154928.1  GI:5524633
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1040)
  AUTHORS   Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R.,
            Russell,P.W. and Orndorff,P.E.
  TITLE     Genetic characterization of Escherichia coli type 1 pilus adhesin
            mutants and identification of a novel binding phenotype
  JOURNAL   J. Bacteriol. 182 (14), 4012-4021 (2000)
  MEDLINE   20327582
   PUBMED   10869080
REFERENCE   2  (bases 1 to 1040)
  AUTHORS   Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R.,
            Russell,P.W. and Orndorff,P.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Microbiology, Pathology, and Parasitology,
            North Carolina State University, College of Veterinary Medicine,
            4700 Hillsborough Street, Raleigh, NC 27606, USA
FEATURES             Location/Qualifiers
     source          1. .1040
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
     gene            47. .949
```

```
                     /gene="fimH"
                     /note="mutant allele"
                     /allele="fimH-205"
     CDS             47. .949
                     /gene="fimH"
                     /note="type 1 pili adhesin"
                     /codon_start=1
                     /transl_table=11
                     /product="FimH precursor"
                     /protein_id="AAD44322.1"
                     /db_xref="GI:5524634"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTRQRGSAYGGVLSNFSGTVKYSGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL
                     SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN
                     YARTGGQVTAGNVQSIIGVTFVYQ"
     variation       282
                     /gene="fimH"
                     /note="confers a leucine to arginine change; results in a
                     temperature sensitive phenotype"
                     /replace="t"
BASE COUNT      252 a    244 c    271 g    273 t
ORIGIN

Alignment Scores:
Pred. No.:                5e-76          Length:          1040
Score:                    828.00         Matches:         157
Percent Similarity:       98.14%         Conservative:    1
Best Local Similarity:    97.52%         Mismatches:      3
Query Match:              97.64%         Indels:          0
DB:                       1              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AF154928 (1-1040)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 244

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db     245 AACGATTATCCGGAAACCATTACAGACTATGTCACACGGCAACGAGGCTCGGCTTATGGC 304

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 364

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 424

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||
Db     425 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 484

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     485 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 544

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 604

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db     605 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 664

Qy     161 Tyr 161
           |||
Db     665 TAT 667

RESULT 20
AF306535
```

```
LOCUS       AF306535                1040 bp    DNA     linear   BCT 21-JUN-2001
DEFINITION  Escherichia coli FimH precursor (fimH) gene, fimH-165 allele,
            complete cds.
ACCESSION   AF306535
VERSION     AF306535.1  GI:11120561
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1040)
  AUTHORS   Harris,S.L., Spears,P.A., Havell,E.A., Hamrick,T.S., Horton,J.R.
            and Orndorff,P.E.
  TITLE     Characterization of Escherichia coli type 1 pilus mutants with
            altered binding specificities
  JOURNAL   J. Bacteriol. 183 (13), 4099-4102 (2001)
  MEDLINE   21289098
   PUBMED   11395476
REFERENCE   2  (bases 1 to 1040)
  AUTHORS   Harris,S.L., Spears,P.A., Havell,E.A., Hamrick,T.S., Horton,J.R.
            and Orndorff,P.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2000) Microbiology, Pathology and Parasitology,
            North Carolina State University, College of Veterinary Medicine,
            4700 Hillsborough Street, Raleigh, NC 27606, USA
FEATURES             Location/Qualifiers
     source          1. .1040
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
     gene            47. .949
                     /gene="fimH"
                     /note="mutant allele"
                     /allele="fimH-165"
     CDS             47. .949
                     /gene="fimH"
                     /note="type 1 pili adhesin; mutation confers a leucine to
                     proline change"
                     /codon_start=1
                     /transl_table=11
                     /product="FimH precursor"
                     /protein_id="AAG30925.1"
                     /db_xref="GI:11120562"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTPQRGSAYGGVLSNFSGTVKYSGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL
                     SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN
                     YARTGGQVTAGNVQSIIGVTFVYQ"
     variation       282
                     /gene="fimH"
                     /replace="g"
BASE COUNT      252 a    245 c    270 g    273 t
ORIGIN

Alignment Scores:
Pred. No.:                6.34e-76       Length:          1040
Score:                    827.00         Matches:         157
Percent Similarity:       98.14%         Conservative:    1
Best Local Similarity:    97.52%         Mismatches:      3
Query Match:              97.52%         Indels:          0
DB:                       1              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AF306535 (1-1040)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 244

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db     245 AACGATTATCCGGAAACCATTACAGACTATGTCACACCGCAACGAGGCTCGGCTTATGGC 304
```

```
Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 364

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 424

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     425 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 484

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     485 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 544

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 604

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     605 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 664

Qy     161 Tyr 161
           |||
Db     665 TAT 667
```

```
RESULT 21
AX354049
LOCUS       AX354049                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 15 from Patent WO0204496.
ACCESSION   AX354049
VERSION     AX354049.1  GI:18618986
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 15 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      191 a    202 c    221 g    223 t
ORIGIN

Alignment Scores:
Pred. No.:                6.26e-76       Length:          837
Score:                    826.00         Matches:         156
Percent Similarity:       98.14%         Conservative:    2
Best Local Similarity:    96.89%         Mismatches:      3
Query Match:              97.41%         Indels:          0
DB:                       6              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354049 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAGAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

```
RESULT 22
AF306536
LOCUS       AF306536                1040 bp    DNA     linear   BCT 21-JUN-2001
DEFINITION  Escherichia coli FimH precursor (fimH) gene, fimH-166 allele,
            complete cds.
ACCESSION   AF306536
VERSION     AF306536.1  GI:11120563
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1040)
  AUTHORS   Harris,S.L., Spears,P.A., Havell,E.A., Hamrick,T.S., Horton,J.R.
            and Orndorff,P.E.
  TITLE     Characterization of Escherichia coli type 1 pilus mutants with
            altered binding specificities
  JOURNAL   J. Bacteriol. 183 (13), 4099-4102 (2001)
  MEDLINE   21289098
   PUBMED   11395476
REFERENCE   2  (bases 1 to 1040)
  AUTHORS   Harris,S.L., Spears,P.A., Havell,E.A., Hamrick,T.S., Horton,J.R.
            and Orndorff,P.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2000) Microbiology, Pathology and Parasitology,
            North Carolina State University, College of Veterinary Medicine,
            4700 Hillsborough Street, Raleigh, NC 27606, USA
FEATURES             Location/Qualifiers
     source          1. .1040
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
     gene            47. .949
                     /gene="fimH"
                     /note="mutant allele"
                     /allele="fimH-166"
     CDS             47. .949
                     /gene="fimH"
                     /note="type 1 pili adhesin; mutation confers a
                     phenylalanine to serine change"
                     /codon_start=1
                     /transl_table=11
                     /product="FimH precursor"
                     /protein_id="AAG30926.1"
                     /db_xref="GI:11120564"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDSQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL
```

```
                     SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN
                     YARTGGQVTAGNVQSIIGVTFVYQ"
     variation       534
                     /gene="fimH"
                     /replace="t"
BASE COUNT      252 a    245 c    270 g    273 t
ORIGIN

Alignment Scores:
Pred. No.:                8.04e-76       Length:          1040
Score:                    826.00         Matches:         157
Percent Similarity:       98.14%         Conservative:    1
Best Local Similarity:    97.52%         Mismatches:      3
Query Match:              97.41%         Indels:          0
DB:                       1              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AF306536 (1-1040)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 244

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 304

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 364

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 424

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     425 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 484

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db     485 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTCCCAGTTTGTG 544

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 604

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     605 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 664

Qy     161 Tyr 161
           |||
Db     665 TAT 667

RESULT 23
AX354039
LOCUS       AX354039                840 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 5 from Patent WO0204496.
ACCESSION   AX354039
VERSION     AX354039.1  GI:18618976
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 5 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      196 a    197 c    222 g    225 t
ORIGIN

Alignment Scores:
Pred. No.:                7.97e-76       Length:          840
Score:                    825.00         Matches:         156
Percent Similarity:       97.52%         Conservative:    1
Best Local Similarity:    96.89%         Mismatches:      4
Query Match:              97.29%         Indels:          0
DB:                       6              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354039 (1-840)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CATGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 24
AX354054
LOCUS       AX354054                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 20 from Patent WO0204496.
ACCESSION   AX354054
VERSION     AX354054.1  GI:18618991
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 20 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      196 a    200 c    220 g    221 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:              1.01e-75        Length:         837
Score:                  824.00          Matches:        156
Percent Similarity:     97.52%          Conservative:   1
Best Local Similarity:  96.89%          Mismatches:     4
Query Match:            97.17%          Indels:         0
DB:                     6               Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AX354054 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAAAAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTAGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 25
AE005662
LOCUS       AE005662               14466 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 281
            of 290.
ACCESSION   AE005662 AE005174
VERSION     AE005662.1  GI:12519313
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 14466)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 14466)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .14466
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .1267
                     /note="O-island #172; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655"
     gene            235. .267
                     /gene="Z5903"
     CDS             235. .267
                     /gene="Z5903"
                     /function="orf; Unknown function"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG59489.1"
                     /db_xref="GI:12519314"
                     /translation="MIDILQVVCQ"
     gene            complement(730. .975)
                     /gene="Z5904"
     CDS             complement(730. .975)
                     /gene="Z5904"
                     /function="orf; Unknown function"
                     /note="Residues 18 to 77 of 81 are 94.99 pct identical to
                     residues 11 to 70 of 196 from GenPept 118 :
                     gi|1787491|gb|AAC74321.1| (AE000222) orf, hypothetical
                     protein [Escherichia coli]"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG59490.1"
                     /db_xref="GI:12519315"
                     /translation="MGAAPVEWLFRQTTQTWGAERYLKDDWHGLQLFAIDGAQFRTPD
                     EPELREYYGSANTSTERQSAYPVMRLVALMNLGITFY"
     gene            complement(1271. .2251)
                     /gene="yjhS"
                     /note="Z5905"
     CDS             complement(1271. .2251)
                     /gene="yjhS"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 326 of 326 are 97.54 pct identical to
                     residues 1 to 326 of 326 from Escherichia coli K-12 Strain
                     MG1655: B4309"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG59491.1"
                     /db_xref="GI:12519316"
                     /translation="MNAIISPDYYYVLTVAGQSNAMAYGEGLPLPDREDAPHPRIKQL
                     ARFAHTHPGGPSCHFNDIIPLTHCPHDVQDMQSYHHPLATNHQTQYGTVGQALHIARK
                     LLPFIPDNAGILIVPCCRGGSAFTAGSEGTYSERHGASHDACRWGTDTPLYQDLVSRT
                     RAALVKNPQNKFLGVCWMQGEFDLMTSDYASHPQHFNHMVEAFRRDLKQYHSQLNNIT
                     DAPWFCGDTTWYWKENFPHAYEAIYGNYQNNILANIIFVDFQQQQGARGLTNAPDEDPD
                     DLSTGYYGSAYRSPENWTTAXRSSHFSSAARRGIISDRFVEAILQFWRER"
     gene            complement(2316. .3530)
                     /gene="yjhT"
                     /note="Z5906"
     CDS             complement(2316. .3530)
                     /gene="yjhT"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 404 of 404 are 98.26 pct identical to
                     residues 1 to 404 of 404 from Escherichia coli K-12 Strain
                     MG1655: B4310"
                     /codon_start=1
                     /transl_table=11
```

/product="orf, hypothetical protein"
/protein_id="AAG59492.1"
/db_xref="GI:12519317"
/translation="MTTWTVRVFTTAEIIYRKTVIALVCHLNCSRQETVTMNKTITAL
AILMASFAANASVLPETPVPFKSGTGAIDNDTVYIGLGSAGTAWYKLETQAKDKKWTA
LAAFPGGPRDQATSAFIDGNLYVFGGIGKNSEGLTQVFNDVHKYNPKTNSWVKLISHA
PMGMAGHVTFVHNGKAYVTGGVNQNIFNGYFEDLNEAGKDSTAVDKINAHYFDKKAED
YFFNKFLLSFDPSTQQWSYAGESPWYGTAGAAVVNKGDKTWLINGEAKPGLRTDAVFE
LDFTGNNLKWNRLAPVSSPDGVAGGFAGISNDSLIFAGGAGFKGSRENYQNGKNYAHE
GLKKSYSTDIHLWHNGKWDKSGELSQGRAYGVSLPWNNSLLIIGGETAGGKAVTDSVL
ISVKDNKVTVQN"

gene complement(3442. .4167)
/gene="yjhA"
/note="Z5907"

CDS complement(3442. .4167)
/gene="yjhA"
/function="orf; Unknown function"
/note="Residues 1 to 241 of 241 are 100.00 pct identical
to residues 1 to 241 of 241 from Escherichia coli K-12
Strain MG1655: B4311"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59493.1"
/db_xref="GI:12519318"
/translation="MKFMKKAKILSGVLLLCFSSPLISQAATLDVRGGYRSGSHAYET
RLKVSEGWQNGWWASMESNTWNTIHDNKKENAALNDVQVEVNYAIKLDDQWTVRPGML
THFSSNGTRYGPYVKLSWDATKDLNFGIRYRYDWKAYRQQDLSGDMSRDNVHRWDGYV
TYHINSDFTFAWQTTLYSKQNDYRYANHKKWATENAFVLQYHMTPDITPYIEYDYLDR
QGVYNGRDNLSENSYRIGVSFKL"

gene 5614. .6216
/gene="fimB"
/note="Z5910"

CDS 5614. .6216
/gene="fimB"
/function="regulator; Cell exterior constituents: Surface
structures"
/note="Residues 1 to 200 of 200 are 100.00 pct identical
to residues 1 to 200 of 200 from Escherichia coli K-12
Strain MG1655: B4312"
/codon_start=1
/transl_table=11
/product="recombinase involved in phase variation;
regulator for fimA"
/protein_id="AAG59494.1"
/db_xref="GI:12519319"
/translation="MKNKADNKKRNFLTHSEIESLLKAANTGPHAARNYCLTLLCFIH
GFRASEICRLRISDIDLKAKCIYIHRLKKGFSTTHPLLNKEVQALKNWLSIRTSYPHA
ESEWVFLSRKGNPLSRQQFYHIISTSGGNAGLSLEIHPHMLRHSCGFALANMGIDTRL
IQDYLGHRNIRHTVWYTASNAGRFYGIWDRARGRQRHAVL"

gene 6694. .7290
/gene="fimE"
/note="Z5911"

CDS 6694. .7290
/gene="fimE"
/function="regulator; Cell exterior constituents: Surface
structures"
/note="Residues 1 to 198 of 198 are 99.49 pct identical to
residues 1 to 198 of 198 from Escherichia coli K-12 Strain
MG1655: B4313"
/codon_start=1
/transl_table=11
/product="recombinase involved in phase variation;
regulator for fimA"
/protein_id="AAG59495.1"
/db_xref="GI:12519320"
/translation="MSKRRYLTGKEVQAMMQAVCYGATGARDYCLILLAYRHGMRISE
LLDLHYQDLDLNEGRINIRRLKNGFSTVHPLRFDEREAVERWTLERANWKGADRTDAI
FISRRGSRLSRQQAYRIIRDAGIEAGTVTQTHPHMLRHACGYELAERGADTRLIQDYL
GHRNIRHTVRYTASNAARFAGLWERNNLINEKLKREEV"

gene 7755. .8303
/gene="fimA"
/note="Z5912"

CDS 7755. .8303
/gene="fimA"
/function="structural component; Cell exterior
constituents: Surface structures"
/note="Residues 1 to 182 of 182 are 96.70 pct identical to
residues 1 to 182 of 182 from Escherichia coli K-12 Strain
MG1655: B4314"
/codon_start=1
/transl_table=11
/product="major type 1 subunit fimbrin (pilin)"
/protein_id="AAG59496.1"
/db_xref="GI:12519321"
/translation="MKIKTLAIVVLSALSLSSTAALAAATTVNGGTVHFKGEVVNAAC
AVDAGSVDQTVQLGQVRTASLAQDGATSSAVGFNIQLNDCDTNVASKAAVAFLGTAID
AGHTNVLALQSSAAGSATNVGVQILDRTGAALTLDGATFSEQTTLNNGTNTIPFQARY
YAIGEATPGAANADATFKVQYQ"

gene 8260. .8907
/gene="fimI"
/note="Z5913"

CDS 8260. .8907
/gene="fimI"
/function="structural component; Cell exterior
constituents: Surface structures"
/note="Residues 1 to 215 of 215 are 98.60 pct identical to
residues 1 to 215 of 215 from Escherichia coli K-12 Strain
MG1655: B4315"
/codon_start=1
/transl_table=11
/product="fimbrial protein"

Alignment Scores:

Pred. No.:	2.67e-74	Length:	14466
Score:	824.00	Matches:	156
Percent Similarity:	97.52%	Conservative:	1
Best Local Similarity:	96.89%	Mismatches:	4
Query Match:	97.17%	Indels:	0
DB:	1	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AE005662 (1-14466)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13586 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 13645

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13646 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 13705

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13706 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 13765

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13766 ACCACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 13825

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db  13826 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 13885

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db  13886 ATTGCCGTGCTTATTTTGCGACAGACCAAAAACTATAACAGCGATGATTTCCAGTTTGTG 13945

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13946 TGGAATATTTACGCCAATAATGATGTGGTAGTGCCTACTGGCGGCTGCGATGTTTCTGCT 14005

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db  14006 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 14065

Qy    161 Tyr 161
```

```
                ||||
Db  14066 TAT 14068

RESULT 26
AP002569
LOCUS       AP002569               168764 bp    DNA     linear   BCT 07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 20/20.
ACCESSION   AP002569 BA000007
VERSION     AP002569.1  GI:13364704
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 168764)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .168764
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .243
                     /gene="ECs5227"
     CDS             79. .243
                     /gene="ECs5227"
                     /note="identical to B4250_ECOLI gi|1790699 (Conserved in
                     E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB38650.1"
                     /db_xref="GI:13364705"
                     /translation="MLSVLLKLLIILVLTPRASTRLLAVRLGYLAVYSMNMSVRKLFR
                     LSIFFVMIVR"
     gene            315. .569
     CDS             315. .569
                     /gene="ECs5228"
                     /note="similar to YJGJ_ECOLI gi|1790700 percent identity
                     92 in 84 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB38651.1"
                     /db_xref="GI:13364706"
                     /translation="MVLEGIHSHDPQARDIAVQYYHAAETAIYDYIARLHPQSAQCVT
                     DFMSTVMSGLSAKAREGHSLEQLCATAALAGEAIKTILKE"
     gene            705. .1166
     CDS             705. .1166
                     /gene="ECs5229"
                     /note="similar to YJGK_ECOLI gi|1790701 percent identity
                     98 in 153 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB38652.1"
                     /db_xref="GI:13364707"
                     /translation="MEKMIIGNIHNLQPWLPQELRQAIEHIKAHVTAETPKGKHDIEG
                     NRLFYLISEDMTEPYEARRAEYHARYLDIQIVLKGQEGMTFSTQPAGTPDTDWLADKD
                     IAFLPEGVDEKTVILNEGDFVVFYPGEVHKPLCAVGAPAQVRKAVVKMLMA"
     gene            1289. .3061
     CDS             1289. .3061
                     /gene="ECs5230"
                     /note="similar to part of (13-616 in 616 aa) YJGL_ECOLI
                     gi|1790702 percent identity 82 in 590 aa"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB38653.1"
                     /db_xref="GI:13364708"
                     /translation="MSKISDLNYSQHITLADNFKQKSEVLNTWRVGMNNFARNAGGQD
                     NTRNILNPKTFLEFLVKIFTLGYVDFSKRSNEAGRNMMAHIESSSYIKNNDGSEIMKF
                     VMNNPEGERADSPKVIIEISLSTITTMGTRQGHTAIIFPQPDGSTNRYERKSFERKDE
                     SSLHLITNKVLACYQREANKEIARLLNNHQKLNNLQKLNNLQKLNNLQKLNNIQKLNN
                     IQELNNSQELNNSQELNNSQELNNSQDLKNSQVSCKGSVDFTITDLLEKSLNNALLAI
                     RNEHLLLMPHVCSESISYLLGENGILEEIDKLYELNDHGIDNDKEGNNEINDIMINLS
                     HILIESLDDAKVNLTPVIHSMLMTFLELPYNNDVKILEWCFNKSMQYFDDSAKIEHAC
                     SVINHINFRRDQSKVAETLFFNLDKEPYKNSPELQELIWKKLVVYVNDFNLSNREKTY
                     LIQRIFNNVESLFNKVPVSILVNDIFMNDFFMKNTEMINWYFPRLLKSYEDEKIYFDK
                     LGYNFNNKESNEEIMKNQPKDVIEEKLNNELKLRFRMMQTILKSEVNVSPFIDQQRLN
                     TLNPPENLRIAIEKFGWKKKTITA"
     gene            complement(3118. .4122)
                     /gene="ECs5231"
     CDS             complement(3118. .4122)
                     /gene="ECs5231"
                     /note="similar to ARGF_ECOLI gi|1786469 percent identity
                     87 in 334 aa, similar to ARGI_ECOLI gi|1790703 percent
                     identity 99 in 334 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
```

```
                     /product="ornithine carbamoyltransferase 1"
                     /protein_id="BAB38654.1"
                     /db_xref="GI:13364709"
                     /translation="MSGFYHKHFLKLLDFTPAELNSLLQLAAKLKADKKSGKEEAKLT
                     GKNIALIFEKDSTRTRCSFEVAAYDQGARVTYLGPSGSQIGHKESIKDTARVLGRMYD
                     GIQYRGYGQEIVETLAEYAGVPVWNGLTNEFHPTQLLADLLTMQEHLPGKPFNEMTLV
                     YAGDARNNMGNSMLEAAALTGLDLRLVAPQACWPEAALVTECRALAQQNGGNITLTED
                     VAKGVEGADFIYTDVWVSMGEAKEKWAERIALLRDYQVNSKMMQLTGNPEVKFLHCLP
                     AFHDDQTTLGKKMAEEFGLHGGMEVTDEVFESAASIVFDQAENRMHTIKAVMVATLSK
                     "
     gene            4284. .4700
                     /gene="ECs5232"
     CDS             4284. .4700
                     /gene="ECs5232"
                     /note="similar to YJGD_ECOLI gi|1790704 percent identity
                     100 in 138 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB38655.1"
                     /db_xref="GI:13364710"
                     /translation="MANPEQLEEQREETRLIIEELLEDGSDPDALYTIEHHLSADDLE
                     TLEKAAVEAFKLGYEVTDPEELEVEDGDIVICCDILSECALNADLIDAQVEQLMTLAE
                     KFDVEYDGWGTYFEDPNGEDGDDEDFVDEDDDGVRH"
     gene            complement(4746. .5039)
                     /gene="ECs5233"
     CDS             complement(4746. .5039)
                     /gene="ECs5233"
                     /note="similar to B4256_ECOLI gi|1790705 percent identity
                     93 in 97 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB38656.1"
                     /db_xref="GI:13364711"
                     /translation="MVGGGGIAPLAGSESDICELQKMYFLPAIRGKGLAKKLALKAME
                     EAREMGFKRCYLETTAFLKEAIGLYEHLGFQHIDYALGCTGHVDCEVRMLREL"
     gene            5442. .6638
                     /gene="ECs5234"
     CDS             5442. .6638
                     /gene="ECs5234"
                     /note="similar to part (5-393 in 393 aa)of YJGN_ECOLI
                     gi|2367367 percent identity 96 in 389 aa (Conserved in
                     E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB38657.1"
                     /db_xref="GI:13364712"
                     /translation="MAQVINEMDVPSHSFVFHGTGERYFLICVVNVLLTIITLGIYLP
                     WALMKCKRYLYANMEVNGQRFSYGITGGNVFFSCLVFVFFYFAILMTVSADMPLIGCV
                     LTLSLLVLLIFMAAKGLRYQALMTSLNGVRFSFNCSMKGVWWVTFFLPILMAIGMGTV
                     FFISTKMLHANSSSSVIVSVVLMAIVGIVSIGIFNGTLYSLVMSFLWSNTSFGIHRFK
                     VKLDTAYCIKYAILAFLALLPFLAVAGYIIFDQILNAYDSSVYANDDIENLQQFMEMQ
                     RKMIIAQLIYYFGIAVSTSYLTVSLRNHFMSNLSLNDGRIRFRSTLTYHGMLYRMCAL
                     VVISGITGGLAYPLLKIWMIDWQAKNTYLLGDLDDLPLINKEEQPDKGFLASISRGIM
```

```
Alignment Scores:
Pred. No.:              4.52e-73        Length:         168764
Score:                  824.00          Matches:        156
Percent Similarity:     97.52%          Conservative:   1
Best Local Similarity:  96.89%          Mismatches:     4
Query Match:            97.17%          Indels:         0
DB:                     1               Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AP002569 (1-168764)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74151 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 74210

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74211 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 74270

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74271 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 74330

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74331 ACCACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 74390

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||::::  ||||||||||||||||||
Db   74391 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 74450

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           |||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db   74451 ATTGCCGTGCTTATTTTGCGACAGACCAAAAACTATAACAGCGATGATTTCCAGTTTGTG 74510

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74511 TGGAATATTTACGCCAATAATGATGTGGTAGTGCCTACTGGCGGCTGCGATGTTTCTGCT 74570

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db   74571 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 74630

Qy     161 Tyr 161
           |||
Db   74631 TAT 74633
```

```
RESULT 27
AF154927
LOCUS       AF154927               1040 bp    DNA     linear   BCT 10-JUL-2000
DEFINITION  Escherichia coli FimH precursor (fimH) gene, fimH-244 allele,
            complete cds.
ACCESSION   AF154927
VERSION     AF154927.1  GI:5524631
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1040)
  AUTHORS   Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R.,
            Russell,P.W. and Orndorff,P.E.
  TITLE     Genetic characterization of Escherichia coli type 1 pilus adhesin
            mutants and identification of a novel binding phenotype
  JOURNAL   J. Bacteriol. 182 (14), 4012-4021 (2000)
  MEDLINE   20327582
   PUBMED   10869080
REFERENCE   2  (bases 1 to 1040)
  AUTHORS   Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R.,
            Russell,P.W. and Orndorff,P.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Microbiology, Pathology, and Parasitology,
            North Carolina State University, College of Veterinary Medicine,
            4700 Hillsborough Street, Raleigh, NC 27606, USA
FEATURES             Location/Qualifiers
     source          1. .1040
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
     gene            47. .949
                     /gene="fimH"
                     /note="mutant allele"
                     /allele="fimH-244"
     CDS             47. .949
                     /gene="fimH"
                     /note="type 1 pili adhesin"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="FimH precursor"
                     /protein_id="AAD44321.1"
                     /db_xref="GI:5524632"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPVVNVGQNLVVDLSTQIFYHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL
                     SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN
                     YARTGGQVTAGNVQSIIGVTFVYQ"
     variation       240
                     /gene="fimH"
                     /note="confers a cysteine to tyrosine change"
                     /replace="g"
BASE COUNT      253 a    244 c    269 g    274 t
ORIGIN

Alignment Scores:
Pred. No.:                1.64e-75       Length:          1040
Score:                    823.00         Matches:         157
Percent Similarity:       98.14%         Conservative:    1
Best Local Similarity:    97.52%         Mismatches:      3
Query Match:              97.05%         Indels:          0
DB:                       1              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AF154927 (1-1040)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db     185 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTACCAT 244

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 304

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 364

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 424

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     425 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 484

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     485 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 544

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 604

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     605 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 664

Qy     161 Tyr 161
           |||
Db     665 TAT 667

RESULT 28
AX354050
LOCUS       AX354050                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 16 from Patent WO0204496.
ACCESSION   AX354050
VERSION     AX354050.1  GI:18618987
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 16 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      195 a    202 c    221 g    219 t
ORIGIN

Alignment Scores:
Pred. No.:                1.62e-75       Length:          837
Score:                    822.00         Matches:         155
Percent Similarity:       98.14%         Conservative:    3
Best Local Similarity:    96.27%         Mismatches:      3
Query Match:              96.93%         Indels:          0
DB:                       6              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354050 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTCATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAGGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATGGCTGTGCTAATTTTGCGACAGACCAATAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCAGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 29
AX354052
LOCUS       AX354052                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 18 from Patent WO0204496.
ACCESSION   AX354052
VERSION     AX354052.1  GI:18618989
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
```

```
  JOURNAL   Patent: WO 0204496-A 18 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      188 a    202 c    226 g    221 t
ORIGIN

Alignment Scores:
Pred. No.:                1.62e-75       Length:          837
Score:                    822.00         Matches:         155
Percent Similarity:       98.14%         Conservative:    3
Best Local Similarity:    96.27%         Mismatches:      3
Query Match:              96.93%         Indels:          0
DB:                       6              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354052 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||||||:::|||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 30
AF288194
LOCUS       AF288194                 937 bp    DNA     linear   BCT 28-MAR-2001
DEFINITION  Escherichia coli type 1 fimbrial adhesin subunit FimH (fimH) gene,
            complete cds.
ACCESSION   AF288194
VERSION     AF288194.1  GI:10946256
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 937)
  AUTHORS   Boudeau,J., Barnich,N. and Darfeuille-Michaud,A.
  TITLE     Type 1 pili-mediated adherence of Escherichia coli strain LF82
            isolated from Crohn's disease is involved in bacterial invasion of
            intestinal epithelial cells
  JOURNAL   Mol. Microbiol. 39 (5), 1272-1284 (2001)
  MEDLINE   21150472
   PUBMED   11251843
REFERENCE   2  (bases 1 to 937)
  AUTHORS   Boudeau,J., Barnich,N. and Darfeuille-Michaud,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2000) Pathogenie Bacterienne Intestinale,
            Laboratoire de Bacteriologie, Faculte de Pharmacie, 28 place Henri
            Dunant, Clermont-Ferrand 63001, France
FEATURES             Location/Qualifiers
     source          1. .937
                     /organism="Escherichia coli"
                     /strain="LF82"
                     /db_xref="taxon:562"
                     /note="isolated from the ileal mucosa of a Crohn's disease
                     patient"
     gene            35. .937
                     /gene="fimH"
     CDS             35. .937
                     /gene="fimH"
                     /codon_start=1
                     /transl_table=11
                     /product="type 1 fimbrial adhesin subunit FimH"
                     /protein_id="AAG24827.1"
                     /db_xref="GI:10946257"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL
                     SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN
                     YARTGGQVTAGNVQSIIGVTFVYQ"
BASE COUNT      217 a    219 c    249 g    252 t
ORIGIN

Alignment Scores:
Pred. No.:                1.84e-75       Length:          937
Score:                    822.00         Matches:         155
Percent Similarity:       98.14%         Conservative:    3
Best Local Similarity:    96.27%         Mismatches:      3
Query Match:              96.93%         Indels:          0
DB:                       1              Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AF288194 (1-937)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 232

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     233 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 292

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||||||:::|||||||||||||||||||||
Db     293 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 352

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     353 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 412

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     413 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 472

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     473 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 532

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     533 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 592

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
```

```
Db    593 CGTGATGTCACCGTTACTTTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 652

Qy    161 Tyr 161
          |||
Db    653 TAT 655

RESULT 31
AX354042
LOCUS       AX354042                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 8 from Patent WO0204496.
ACCESSION   AX354042
VERSION     AX354042.1  GI:18618979
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 8 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      192 a    200 c    225 g    220 t
ORIGIN

Alignment Scores:
Pred. No.:                2.6e-75       Length:          837
Score:                    820.00        Matches:         155
Percent Similarity:       96.89%        Conservative:    1
Best Local Similarity:    96.27%        Mismatches:      5
Query Match:              96.70%        Indels:          0
DB:                       6             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354042 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   |||||||||||||||||||||||||||||||||||||||||   |||||||||
Db     76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAACCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CATGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 32
AX354045
LOCUS       AX354045                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 11 from Patent WO0204496.
ACCESSION   AX354045
VERSION     AX354045.1  GI:18618982
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 11 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      193 a    202 c    224 g    218 t
ORIGIN

Alignment Scores:
Pred. No.:                3.29e-75      Length:          837
Score:                    819.00        Matches:         154
Percent Similarity:       98.14%        Conservative:    4
Best Local Similarity:    95.65%        Mismatches:      3
Query Match:              96.58%        Indels:          0
DB:                       6             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354045 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db    136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||:::|||||||||||||||||||||||:::|||||||||||||||||||||
Db    196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 33
AX354056
```

```
LOCUS       AX354056                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 22 from Patent WO0204496.
ACCESSION   AX354056
VERSION     AX354056.1  GI:18618993
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 22 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      193 a    202 c    224 g    218 t
ORIGIN
```

Alignment Scores:

Pred. No.:	3.29e-75	Length:	837
Score:	819.00	Matches:	154
Percent Similarity:	98.14%	Conservative:	4
Best Local Similarity:	95.65%	Mismatches:	3
Query Match:	96.58%	Indels:	0
DB:	6	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AX354056 (1-837)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||:::||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

```
RESULT 34
AX354038
LOCUS       AX354038                840 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 4 from Patent WO0204496.
ACCESSION   AX354038
VERSION     AX354038.1  GI:18618975
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 4 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      195 a    202 c    224 g    219 t
ORIGIN
```

Alignment Scores:

Pred. No.:	3.3e-75	Length:	840
Score:	819.00	Matches:	154
Percent Similarity:	98.14%	Conservative:	4
Best Local Similarity:	95.65%	Mismatches:	3
Query Match:	96.58%	Indels:	0
DB:	6	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AX354038 (1-840)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||:::||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

```
RESULT 35
AF089840
LOCUS       AF089840                982 bp    DNA     linear   BCT 16-SEP-1998
DEFINITION  Escherichia coli FimH (fimH) gene, complete cds.
ACCESSION   AF089840
VERSION     AF089840.1  GI:3599570
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
```

```
            Escherichia.
REFERENCE   1  (bases 1 to 982)
  AUTHORS   Pouttu,R. and Korhonen,T.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1998) Department of Biosciences, Division of
            General Microbiology, University of Helsinki, Viikinkaari 9 C,
            P.O.Box 56, Helsinki FIN-00014, Finland
FEATURES             Location/Qualifiers
     source          1. .982
                     /organism="Escherichia coli"
                     /strain="IHE3034"
                     /serotype="O18:K1:H7"
                     /db_xref="taxon:562"
                     /note="meningitis isolate"
     gene            47. .949
                     /gene="fimH"
     CDS             47. .949
                     /gene="fimH"
                     /codon_start=1
                     /transl_table=11
                     /product="FimH"
                     /protein_id="AAC35864.1"
                     /db_xref="GI:3599571"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYNGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL
                     SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN
                     YARTGGQVTAGNVQSIIGVTFVYQ"
BASE COUNT      236 a    232 c    257 g    257 t
ORIGIN

Alignment Scores:
Pred. No.:                3.95e-75        Length:          982
Score:                    819.00          Matches:         154
Percent Similarity:       98.14%          Conservative:    4
Best Local Similarity:    95.65%          Mismatches:      3
Query Match:              96.58%          Indels:          0
DB:                       1               Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AF089840 (1-982)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 244

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db     245 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 304

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||||||:::|||||||||||||||||||||
Db     305 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 364

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 424

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     425 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 484

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     485 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 544

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 604

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     605 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 664

Qy     161 Tyr 161
           |||
Db     665 TAT 667

RESULT 36
AF317710
LOCUS       AF317710               1008 bp    DNA     linear   BCT 31-JAN-2001
DEFINITION  Escherichia coli FimH (fimH) gene, complete cds.
ACCESSION   AF317710
VERSION     AF317710.1  GI:12620397
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1008)
  AUTHORS   Elliott,S.J. and Kim,K.S.
  TITLE     Virulence factors of meningitic Escherichia coli
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1008)
  AUTHORS   Elliott,S.J. and Kim,K.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) Pediatric Infectious Disease, Johns Hopkins
            Medical School, Ross 1109/ 720 Rutland Rd, Baltimore, MD 21205, USA
FEATURES             Location/Qualifiers
     source          1. .1008
                     /organism="Escherichia coli"
                     /strain="RS218"
                     /serotype="O18:H7:K1"
                     /db_xref="taxon:562"
                     /note="biovar: meningitis"
     gene            44. .955
                     /gene="fimH"
     CDS             44. .955
                     /gene="fimH"
                     /note="adhesin"
                     /codon_start=1
                     /transl_table=11
                     /product="FimH"
                     /protein_id="AAG60686.1"
                     /db_xref="GI:12620398"
                     /translation="MIVMKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANV
                     YVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSGTVKYN
                     GSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNN
                     YNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLG
                     YYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGL
                     TANYARTGGQVTAGNVQSIIGVTFVYQ"
BASE COUNT      246 a    236 c    263 g    263 t
ORIGIN

Alignment Scores:
Pred. No.:                4.08e-75        Length:          1008
Score:                    819.00          Matches:         154
Percent Similarity:       98.14%          Conservative:    4
Best Local Similarity:    95.65%          Mismatches:      3
Query Match:              96.58%          Indels:          0
DB:                       1               Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AF317710 (1-1008)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 250

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||||
Db     251 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 310

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||||||:::|||||||||||||||||||||
Db     311 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 370
```

```
Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     371 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 430

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     431 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 490

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     491 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 550

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     551 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 610

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     611 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 670

Qy     161 Tyr 161
           |||
Db     671 TAT 673
```

```
RESULT 37
AF306537
LOCUS       AF306537               1040 bp    DNA     linear   BCT 08-NOV-2000
DEFINITION  Escherichia coli FimH precursor (fimH) gene, fimH-171 allele,
            complete cds.
ACCESSION   AF306537
VERSION     AF306537.1  GI:11120565
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1040)
  AUTHORS   Harris,S.L., Spears,P.A., Havell,E.A., Hamrick,T.S., Horton,J.R.
            and Orndorff,P.E.
  TITLE     Isolation and characterization of Escherichia coli type 1 pili
            mutants that have altered binding specificities
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1040)
  AUTHORS   Harris,S.L., Spears,P.A., Havell,E.A., Hamrick,T.S., Horton,J.R.
            and Orndorff,P.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2000) Microbiology, Pathology and Parasitology,
            North Carolina State University, College of Veterinary Medicine,
            4700 Hillsborough Street, Raleigh, NC 27606, USA
FEATURES             Location/Qualifiers
     source          1. .1040
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
     gene            47. .547
                     /gene="fimH"
                     /note="mutant allele"
                     /allele="fimH-171"
     CDS             47. .547
                     /gene="fimH"
                     /note="type 1 pili adhesin; mutation confers a tryptophan
                     to nonsense mutation; truncated protein"
                     /codon_start=1
                     /transl_table=11
                     /product="FimH precursor"
                     /protein_id="AAG30927.1"
                     /db_xref="GI:11120566"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDFQFV"
     variation       546
                     /gene="fimH"
                     /replace="g"
BASE COUNT      253 a    244 c    269 g    274 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                4.22e-75        Length:          1040
Score:                    819.00          Matches:         157
Percent Similarity:       98.14%          Conservative:    1
Best Local Similarity:    97.52%          Mismatches:      3
Query Match:              96.58%          Indels:          0
DB:                       1               Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AF306537 (1-1040)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 244

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 304

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     305 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 364

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     365 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 424

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     425 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 484

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     485 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 544

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     545 TAGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 604

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     605 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 664

Qy     161 Tyr 161
           |||
Db     665 TAT 667
```

```
RESULT 38
AX354043
LOCUS       AX354043               837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 9 from Patent WO0204496.
ACCESSION   AX354043
VERSION     AX354043.1  GI:18618980
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 9 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      194 a    200 c    223 g    220 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:              4.17e-75        Length:         837
Score:                  818.00          Matches:        154
Percent Similarity:     97.52%          Conservative:   3
Best Local Similarity:  95.65%          Mismatches:     4
Query Match:            96.46%          Indels:         0
DB:                     6               Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AX354043 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||:::|||||||||||||||||:::||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGCTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTTTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 39
AX354035.
LOCUS       AX354035                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 1 from Patent WO0204496.
ACCESSION   AX354035
VERSION     AX354035.1  GI:18618972
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 1 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      186 a    204 c    228 g    219 t
ORIGIN

Alignment Scores:
Pred. No.:              5.29e-75        Length:         837
Score:                  817.00          Matches:        155
Percent Similarity:     96.89%          Conservative:   1
Best Local Similarity:  96.27%          Mismatches:     5
Query Match:            96.34%          Indels:         0
DB:                     6               Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AX354035 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTCCGGGATCGTAAAATATAGTGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCAGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGGTTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGCTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 40
AX354047
LOCUS       AX354047                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 13 from Patent WO0204496.
ACCESSION   AX354047
VERSION     AX354047.1  GI:18618984
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 13 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      195 a    200 c    222 g    220 t
ORIGIN

Alignment Scores:
Pred. No.:              6.7e-75         Length:         837
Score:                  816.00          Matches:        154
Percent Similarity:     97.52%          Conservative:   3
Best Local Similarity:  95.65%          Mismatches:     4
Query Match:            96.23%          Indels:         0
DB:                     6               Gaps:           0
```

```
US-09-900-575-29_COPY_26_186 (1-161) x AX354047 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
             ||||||||||:::|||||||||||||||||||:::|||||||||||||||||||||||
Db    196 AGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCCGGTGAGCAGTGCGGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 41
AX354048
LOCUS       AX354048                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 14 from Patent WO0204496.
ACCESSION   AX354048
VERSION     AX354048.1  GI:18618985
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 14 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      195 a    200 c    222 g    220 t
ORIGIN

Alignment Scores:
Pred. No.:              6.7e-75        Length:         837
Score:                  816.00         Matches:        154
Percent Similarity:     97.52%         Conservative:   3
Best Local Similarity:  95.65%         Mismatches:     4
Query Match:            96.23%         Indels:         0
DB:                     6              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x AX354048 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
             ||||||||||:::|||||||||||||||||||:::|||||||||||||||||||||||
Db    196 AGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 42
ECFIM
LOCUS       ECFIM                   3356 bp    DNA     linear   BCT 29-MAR-1999
DEFINITION  Escherichia coli fimD, fimF, fimG, fimH, uxaA & gntP genes.
ACCESSION   AJ225176
VERSION     AJ225176.1  GI:3286745
KEYWORDS    fimbrial subunit; fimD gene; fimF gene; fimG gene; fimH gene; gntP
            gene; uxaA gene.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 3356)
  AUTHORS   Klemm,P. and Christiansen,G.
  TITLE     Three fim genes required for the regulation of length and mediation
            of adhesion of Escherichia coli type 1 fimbriae
  JOURNAL   Mol. Gen. Genet. 208 (3), 439-445 (1987)
  MEDLINE   88038337
   PUBMED   2890081
REFERENCE   2  (bases 1 to 3356)
  AUTHORS   Klemm,P., Tong,S., Nielsen,H. and Conway,T.
  TITLE     The gntP gene of Escherichia coli involved in gluconate uptake
  JOURNAL   J. Bacteriol. 178 (1), 61-67 (1996)
  MEDLINE   96125229
   PUBMED   8550444
REFERENCE   3  (bases 1 to 3356)
  AUTHORS   Marc,D. and Dho-Moulin,M.
  TITLE     Analysis of the fim cluster of an avian O2 strain of Escherichia
            coli: serogroup-specific sites within fimA and nucleotide sequence
            of fimI
  JOURNAL   J. Med. Microbiol. 44 (6), 444-452 (1996)
  MEDLINE   96235989
   PUBMED   8636962
REFERENCE   4  (bases 1 to 3356)
  AUTHORS   Blattner,F., Plunkett,G., Bloch,C., Perna,N., Burland,V., Riley,M.,
            Collado-Vides,J., Glasner,J., Rode,C., Mayhew,G., Gregor,J.,
            Davis,N., Kirkpatrick,H., Goeden,M., Rose,D., Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
```

```
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   5  (bases 1 to 3356)
  AUTHORS   Marc,D., Arne,P., Bree,A. and Dho-Moulin,M.
  TITLE     Colonization ability and pathogenic properties of a fim- mutant of
            an avian strain of Escherichia coli
  JOURNAL   Res. Microbiol. 149 (7), 473-485 (1998)
  MEDLINE   98439336
   PUBMED   9766199
REFERENCE   6  (bases 1 to 3356)
  AUTHORS   Marc,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1997) Station de Pathologie Aviaire et
            Parasitologie, Institut National de la Recherche Agronomique, INRA
            - Centre de Tours, Nouzilly 37380, France
FEATURES             Location/Qualifiers
     source          1. .3356
                     /organism="Escherichia coli"
                     /strain="MT78"
                     /db_xref="taxon:562"
                     /note="avian pathogenic strain of O2 serogroup (citation
                     [2]"
     gene            1. .264
                     /gene="fimD"
     CDS             <1. .264
                     /gene="fimD"
                     /codon_start=1
                     /transl_table=11
                     /product="FimD protein"
                     /protein_id="CAA12420.1"
                     /db_xref="GI:3286746"
                     /translation="ARVGIKLLMTLTHNNKPLPFGAMVTSESSQSSGIVADNGQVYLS
                     GMPLAGKVQVKWGEEENAHCVANYQLPPESQQQLLTQLSAECR"
     gene            274. .804
                     /gene="fimF"
     CDS             274. .804
                     /gene="fimF"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="minor fimbrial subunit, precursor polypeptide"
                     /protein_id="CAA12421.1"
                     /db_xref="GI:3286747"
                     /db_xref="SPTREMBL:O87737"
                     /translation="MRNKPFYLLCAFLWLAVSHALAADSTITIRGYVRDNGCSVAAES
                     TNFTVDLMENAAKQFNNIGATTPVVPFRILLSPCGNAVSAVKVGFTGVADSHNANLLA
                     LENTVSAAAGLGIQLLNEQQNEIPLNAPSSAISWTTLTPGKPNTLNFYARLMATQVPV
                     TAGHINATATFTLEYQ"
     gene            817. .1320
                     /gene="fimG"
     CDS             817. .1320
                     /gene="fimG"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="minor fimbrial subunit, precursor polypeptide"
                     /protein_id="CAA12422.1"
                     /db_xref="GI:3286748"
                     /db_xref="SPTREMBL:O87738"
                     /translation="MKWCKRGYLLAAMLAFASATIQAADVTITVNGKVVAKPCTVSTT
                     NATVDLGDLYSFSLMSAGAASAWHDVALELTNCPVGTSRVTASFSGAADSTGYYKNQG
                     TAQNIQLELQDDSGNTLNTGATKTVQVDDSSQSAHFPLQVRALTVNGGATQGTIQAVI
                     SITYTYS"
     gene            1340. .2242
                     /gene="fimH"
     CDS             1340. .2242
                     /gene="fimH"
                     /codon_start=1
                     /transl_table=11
                     /product="type 1 fimbriae adhesin, precursor polypeptide"
                     /protein_id="CAA12423.1"
                     /db_xref="GI:3286749"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPAVNVGQKLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSS
                     YPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS
                     DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL
                     SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTAN
                     YARTGGQVTAGNVQSIIGVTFVYQ"
     gene            2415. .2631
                     /gene="gntP"
     misc_feature    2415. .2631
                     /gene="gntP"
                     /note="3' and 5' regions of gene gntP (most of the gene
                     deleted between nt 2523 and 2568)"
                     /pseudo
     gene            2971. .3356
                     /gene="uxaA"
     CDS             2971. .>3356
                     /gene="uxaA"
                     /citation=[4]
                     /codon_start=1
                     /transl_table=11
                     /product="UxuA protein"
                     /protein_id="CAA12424.1"
                     /db_xref="GI:3286750"
                     /db_xref="SPTREMBL:O87739"
                     /translation="MEQTWRWYGPNDPVSLADVRQAGATGVVTALHHIPNGEVWSVEE
                     ILKRKAIVEDAGLVWSVVESVPIHEDIKTHTGNYEQWIANYQQTLRNLAQCGIRTVCY
                     NFMPVLDWTRTDLEYVLPDGSKALRF"
BASE COUNT      845 a    819 c    866 g    826 t
ORIGIN

Alignment Scores:
Pred. No.:                3.31e-74        Length:          3356
Score:                    816.00          Matches:         154
Percent Similarity:       97.52%          Conservative:    3
Best Local Similarity:    95.65%          Mismatches:      4
Query Match:              96.23%          Indels:          0
DB:                       1               Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x ECFIM (1-3356)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
             |||   ||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db    1478 CCTGCCGTGAATGTGGGGCAAAAGCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 1537

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1538 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 1597

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||||||||:::|||||||||||||||||||||
Db    1598 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 1657

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1658 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 1717

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    1718 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 1777

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1778 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 1837

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1838 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 1897

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    1898 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 1957

Qy     161 Tyr 161
           |||
```

```
Db   1958 TAT 1960

RESULT 43
AX354036
LOCUS       AX354036                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 2 from Patent WO0204496.
ACCESSION   AX354036
VERSION     AX354036.1  GI:18618973
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 2 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      191 a    202 c    224 g    220 t
ORIGIN

Alignment Scores:
Pred. No.:                 1.73e-74      Length:          837
Score:                     812.00        Matches:         153
Percent Similarity:        96.89%        Conservative:    3
Best Local Similarity:     95.03%        Mismatches:      5
Query Match:               95.75%        Indels:          0
DB:                        6             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354036 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::||||||||||   ||||||||||:::||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGATCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCAGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGCTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 44
AX354053
```

```
LOCUS       AX354053                837 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 19 from Patent WO0204496.
ACCESSION   AX354053
VERSION     AX354053.1  GI:18618990
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Langermann,S., Revel,A., Auguste,C. and Burlein,J.
  TITLE     Fimh adhesin proteins and methods of use
  JOURNAL   Patent: WO 0204496-A 19 17-JAN-2002;
            MEDIMMUNE, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      194 a    199 c    223 g    221 t
ORIGIN

Alignment Scores:
Pred. No.:                 1.73e-74      Length:          837
Score:                     812.00        Matches:         154
Percent Similarity:        97.52%        Conservative:    3
Best Local Similarity:     95.65%        Mismatches:      4
Query Match:               95.75%        Indels:          0
DB:                        6             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x AX354053 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::||||||||||||||||||||||:::||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||   ||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCTGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 45
AF154929
LOCUS       AF154929                1039 bp    DNA     linear   BCT 10-JUL-2000
DEFINITION  Escherichia coli FimH precursor (fimH) gene, fimH-208 allele,
            complete cds.
ACCESSION   AF154929
```

```
VERSION     AF154929.1  GI:5524635
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1039)
  AUTHORS   Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R.,
            Russell,P.W. and Orndorff,P.E.
  TITLE     Genetic characterization of Escherichia coli type 1 pilus adhesin
            mutants and identification of a novel binding phenotype
  JOURNAL   J. Bacteriol. 182 (14), 4012-4021 (2000)
  MEDLINE   20327582
   PUBMED   10869080
REFERENCE   2  (bases 1 to 1039)
  AUTHORS   Hamrick,T.S., Harris,S.L., Spears,P.A., Havell,E.A., Horton,J.R.,
            Russell,P.W. and Orndorff,P.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Microbiology, Pathology, and Parasitology,
            North Carolina State University, College of Veterinary Medicine,
            4700 Hillsborough Street, Raleigh, NC 27606, USA
FEATURES             Location/Qualifiers
     source          1. .1039
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
     gene            47. .436
                     /gene="fimH"
                     /note="mutant allele"
                     /allele="fimH-208"
     CDS             47. .436
                     /gene="fimH"
                     /codon_start=1
                     /transl_table=11
                     /product="FimH precursor"
                     /protein_id="AAD44323.1"
                     /db_xref="GI:5524636"
                     /translation="MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVN
                     LAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVNIVAVA
                     IHFLPPAKRRALFIIRERISRGRWRFI"
     variation       337
                     /gene="fimH"
                     /note="deletion creates in a frameshift; results in a
                     truncated protein"
BASE COUNT      251 a    244 c    270 g    274 t
ORIGIN
```

Alignment Scores:

Pred. No.:	7.26e-74	Length:	1039
Score:	807.00	Matches:	157
Percent Similarity:	98.14%	Conservative:	1
Best Local Similarity:	97.52%	Mismatches:	3
Query Match:	95.17%	Indels:	1
DB:	1	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x AF154929 (1-1039)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     185 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 244

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 304

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           |||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db     305 GGCGTGTTATCTAATTTTTCCGGGACCGT-AAATATAGTGGCAGTAGCTATCCATTTCCT 363

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     364 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 423

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     424 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 483

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     484 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 543

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     544 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 603

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     604 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 663

Qy     161 Tyr 161
           |||
Db     664 TAT 666
```

Search completed: November 28, 2002, 19:59:12
Job time : 2666 secs

NO

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2002, 19:09:21 ; Search time 49 Seconds
(without alignments)
1007.652 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09900575/runat_22112002_130709_4580/app_query.fasta_1.327
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=45
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09900575_@CGN_1_1_17_@runat_22112002_130709_4580 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		96.5	11.4	30549	4	US-09-134-001C-322	Sequence 322, App
c 2		84.5	10.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
3		79	9.3	2717	1	US-07-723-002C-1	Sequence 1, Appli
4		78	9.2	537	4	US-09-134-001C-2157	Sequence 2157, Ap
5		77.5	9.1	1777	1	US-08-229-781-54	Sequence 54, Appl
6		77.5	9.1	1777	1	US-08-630-918-54	Sequence 54, Appl
7		77.5	9.1	1777	4	US-09-004-422-54	Sequence 54, Appl
c 8		77.5	9.1	2002	1	US-08-309-341-3	Sequence 3, Appli
c 9		77.5	9.1	2002	1	US-08-608-267-3	Sequence 3, Appli
c 10		77.5	9.1	2002	1	US-08-608-452-3	Sequence 3, Appli
c 11		77.5	9.1	2002	1	US-08-608-224-3	Sequence 3, Appli
c 12		77.5	9.1	2002	2	US-08-967-149-3	Sequence 3, Appli
13		75.5	8.9	1916	3	US-08-714-918-96	Sequence 96, Appl
14		75.5	8.9	1916	4	US-09-265-315-96	Sequence 96, Appl
15		75.5	8.9	1916	4	US-09-265-315-96	Sequence 96, Appl
16		75.5	8.9	1916	4	US-09-266-417-96	Sequence 96, Appl
17		75	8.8	420	4	US-09-641-638-65	Sequence 65, Appl
18		75	8.8	957	1	US-08-684-862-11	Sequence 11, Appl
19		75	8.8	5319	1	US-08-169-927-1	Sequence 1, Appli
20		74.5	8.8	7886	4	US-09-453-702B-102	Sequence 102, App
c 21		74	8.7	17056	4	US-09-245-041-3	Sequence 3, Appli
22		73	8.6	4162	4	US-09-453-702B-259	Sequence 259, App
23		72.5	8.5	1636	3	US-08-908-643C-1	Sequence 1, Appli
24		72	8.5	3531	2	US-08-714-402-1	Sequence 1, Appli
25		72	8.5	3698	4	US-09-327-536-1	Sequence 1, Appli
26		72	8.5	25165	4	US-09-453-702B-39	Sequence 39, Appl
27		71.5	8.4	1843	1	US-07-918-023-1	Sequence 1, Appli
c 28		71.5	8.4	8855	2	US-08-542-003-1	Sequence 1, Appli
c 29		71.5	8.4	8855	2	US-08-322-760A-1	Sequence 1, Appli
c 30		71.5	8.4	8855	4	US-09-236-949-1	Sequence 1, Appli
c 31		71.5	8.4	9100	2	US-08-743-637B-27	Sequence 27, Appl
c 32		71.5	8.4	9100	3	US-08-526-840B-27	Sequence 27, Appl
c 33		71.5	8.4	9421	2	US-08-370-319C-2	Sequence 2, Appli
c 34		71.5	8.4	9421	4	US-09-224-834-2	Sequence 2, Appli
35		71	8.4	786	4	US-09-110-959A-1	Sequence 1, Appli
c 36		71	8.4	1463	4	US-09-071-035-263	Sequence 263, App
c 37		71	8.4	4977	4	US-09-071-035-257	Sequence 257, App
c 38		71	8.4	4977	4	US-09-071-035-261	Sequence 261, App
c 39		71	8.4	4977	4	US-09-071-035-265	Sequence 265, App
40		70.5	8.3	32768	4	US-08-961-527-71	Sequence 71, Appl
41		70	8.3	1870	3	US-09-295-186-9	Sequence 9, Appli
c 42		69.5	8.2	9319	4	US-08-976-259-85	Sequence 85, Appl
43		69.5	8.2	22671	4	US-08-976-259-14	Sequence 14, Appl
44		69	8.1	1408	2	US-08-712-072C-1	Sequence 1, Appli
45		69	8.1	1860	4	US-09-620-412C-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-09-134-001C-322
; Sequence 322, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 322
; LENGTH: 30549
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-322

Alignment Scores:

Pred. No.:	1.21	Length:	30549
Score:	96.50	Matches:	49
Percent Similarity:	45.00%	Conservative:	32
Best Local Similarity:	27.22%	Mismatches:	62
Query Match:	11.38%	Indels:	37
DB:	4	Gaps:	11

US-09-900-575-29_COPY_26_186 (1-161) x US-09-134-001C-322 (1-30549)

```
Qy       2 ValValAsnValGlyGlnAsnLeuValValAspLeu---------------------- 13
           |||   ||||||   ||||||      |||   |||
Db    5770 GTTCGAAATGTTCCACAAAATGCACAAGTCACTCTATATCAATCAAATGGAACTGTTATT 5829
```

```
Qy      14 ---SerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyrValThr 32
Db    5830 CCTAATACAAATACAACTATAGATTCTAATGGTATAGCTACTGTAACAATTCAAGGCACT 5889

Qy      33 LeuGlnArgGly------SerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrVal 50
Db    5890 CTACCAACCGGAAATATTACTGCTAAAACCTCAATGACAAATAATGTAACGTACACT--- 5946

Qy      51 LysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyr 70
Db    5947 AAACAAAATAGTAGTGGAATTGCTTCAAATACAACTGAAGATATAAGT------GTTTTT 6000

Qy      71 AsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAla 90
Db    6001 TCAGAAAACAGTGATCAA----------------------------GTAAATGTTACC 6030

Qy      91 GlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsn 110
Db    6031 GCTGGCATGCAAGCTAAAAATGAT---------GGTATTAAAATAATTAAAGGTACA--- 6078

Qy     111 AsnTyrAsnSerAspAspPhe---GlnPheValTrpAsnIleTyrAlaAsnAsnAspVal 129
Db    6079 AACTATAATTTTAATGACTTCAATAGTTTCATAAGTAATATACCAGCCCATTCTACTCTT 6138

Qy     130 Val---------ValProThr---------GlyGlyCysAspValSerAlaArgAspVal 143
Db    6139 ACATGGAACGAGGAGCCTAATAGTTGGAAAAACAACATCGGTACTACAACAAAAACTGTT 6198

Qy     144 ThrValThrLeuProAspTyrArgGly------SerValProIleProLeuThrValTyr 161
Db    6199 ACAGTTACTCTACCTAATCATCAAGGTACGAGAACTGTAGATATTCCAATAACAATCTAT 6258

RESULT 2
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:              5.75e+04        Length:         4403765
Score:                  84.50           Matches:        42
Percent Similarity:     37.76%          Conservative:   12
Best Local Similarity:  29.37%          Mismatches:     72
Query Match:            9.96%           Indels:         17
DB:                     4               Gaps:           5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-103-840A-2 (1-4403765)

Qy      27 IleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPhe 46
Db 586389 ATCTCCAGCAACTCGACGAGGAAGTCGGGGCCCAGTTTGCGCACCAGCGCCTCGACCACC 586330

Qy      47 SerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrPro 66

Db 586329 GCGGGATGGGTGAGCCCCTCGGGGTCACGGACCACACTGCCGCTGCCGTGCCGCACCTCG 586270

Qy      67 ---ArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThr 85
Db 586269 ATCAGGCCCATCTGTTGCAGTCGCGCC---AGACCCTGGCGTAGCGAGGTGCGGTTGACA 586213

Qy      86 ProValSerSerAla------------GlyGlyLeuValIleLysAlaGly--------- 98
Db 586212 CCGAGCCGCTCTGCCAGGTCTCGCTCGGGCGGCAGGGTCGAGCCCGGCGGGAAGACGCCG 586153

Qy      99 ------------SerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSer 114
Db 586152 TCGAGGATGGCGTCGGCGATCGTCGTCGCGATCCGCTCATCCACCCGCTGCCGATCCGGC 586093

Qy     115 AspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspVal------------Val 130
Db 586092 GGCTGAAAGACACTTGACTGGTTCATTGGCTCAACCAATATAGTGGGCGAATCGACAACT 586033

Qy     131 ValProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAspTyr 150
Db 586032 CAGCCAGCAGGAGGTGCCGCAGTGGACGATGAAACGCGGTCGGCGGCCTTGCCGCGGTGG 585973

Qy     151 ArgGlySer 153
Db 585972 CGCGGTGCA 585964

RESULT 3
US-07-723-002C-1
; Sequence 1, Application US/07723002C
; Patent No. 5447862
;  GENERAL INFORMATION:
;    APPLICANT:  Heim, Jutta
;    APPLICANT:  Meyhack, Bernd
;    APPLICANT:  Gysler, Christof
;    APPLICANT:  Visser, Jacob
;    APPLICANT:  Kester, Hermanus Cornelis Maria
;    TITLE OF INVENTION:  No. 5447862el Expression System
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/723,002C
;      FILING DATE:  28-JUN-1991
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8702475
;      FILING DATE:  04-FEB-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 150,880
;      FILING DATE:  29-JAN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8818046.8
;      FILING DATE:  28-JUL-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 8914666.6
;      FILING DATE:  26-JUN-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 384,898
;      FILING DATE:  24-JUL-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Foley, Shawn P.
;      REGISTRATION NUMBER:  33,071
```

```
;      REFERENCE/DOCKET NUMBER:  4-16317/+/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919)541-8615
;      TELEFAX:  (919)541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2717 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger N756
;    IMMEDIATE SOURCE:
;      CLONE:  Plasmid pCG3B11 (DSM 3916)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(689..889, 955..1098, 1161..1286, 1350..1445,
;      LOCATION:  1503..2054)
;      OTHER INFORMATION:  /transl_except= (pos: 1096 .. 1098, aa: Tyr)
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  689..746
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  747..889
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  890..954
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  955..1097
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1098..1159
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1160..1286
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1287..1349
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1350..1445
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1446..1502
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1503..2054
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..688
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  2058..2717
US-07-723-002C-1

Alignment Scores:
Pred. No.:                5.42          Length:          2717
Score:                    79.00         Matches:         38
Percent Similarity:       39.13%        Conservative:    25
Best Local Similarity:    23.60%        Mismatches:      72
Query Match:              9.32%         Indels:          26
DB:                       1             Gaps:            7

US-09-900-575-29_COPY_26_186 (1-161) x US-07-723-002C-1 (1-2717)

Qy       9 LeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr 28
           |||      |||      :::::::::          |||:::|||           :::
Db    1533 CTCGGTACCGACGCCGACAGCCGTGTCTCCATCACCAACAACTACATCAACGGCGAGTCT 1592

Qy      29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyVal---LeuSerAsnPheSer 47
           ::::|||      ||||||::::::   |||||||||      :::|||   |||
Db    1593 GACTACTCTGCTACTTGCGACGGCCACCACTACTGGAACGTGTACCTTGACGGCTCTAGC 1652

Qy      48 GlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArg 67
           |||   :::|||||||   :::|||   :::   |||:::          |||:::
Db    1653 GACAAGGTCACCTTCAGTGGC---AACTACCTGTACAAGACCTCCGGCCGTGCCCCCAAG 1709

Qy      68 Val--------------------ValTyrAsnSerArgThrAspLysProTrpProVal 80
           |||                    :::|||||:::      :::
Db    1710 GTCCAGGACAACACCTACCTCCACATCTACAACAACTACTGGGAGAACAACTCGGGCCAC 1769

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           |||      :::              :::|||||   |||:::      |||:::
Db    1770 GCTTTCGAGATC------------GGCTCCGGTGGCTACGTCCTCGCCGAGGGTAACTAC 1817

Qy     101 IleAla---ValLeuIleLeuArgGlnThrAsnAsn--------------TyrAsnSer 114
           :::   |||      :::   :::|||:::                    :::::|||
Db    1818 TTCTCCAACGTCGACACCGTCCTCGAGACCGACACCTTCGAGGGTGCTCTCTTCTCCTCT 1877

Qy     115 AspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGly 134
           |||                  |||                 |||
Db    1878 GACAGCGCCTCCTCCACCTGCGAGTCCTACATTGGCCGTTCCTGCGTTGCCAACGTCAAC 1937

Qy     135 GlyCysAspValSerAlaArgAspValThrVal--------------------ThrLeu 147
           |||   |||::::::          ||||||                    ||||||
Db    1938 GGCGGTGACCTCACCGGCACCTCCACCACCGTCCTCTCCAACCTCAGCGGCGACACCCTC 1997

Qy     148 Pro 148
           |||
Db    1998 CCC 2000

RESULT 4
US-09-134-001C-2157
; Sequence 2157, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2157
;   LENGTH: 537
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2157

Alignment Scores:
Pred. No.:                0.663         Length:          537
Score:                    78.00         Matches:         20
Percent Similarity:       48.44%        Conservative:    11
Best Local Similarity:    31.25%        Mismatches:      19
Query Match:              9.20%         Indels:          14
DB:                       4             Gaps:            2

US-09-900-575-29_COPY_26_186 (1-161) x US-09-134-001C-2157 (1-537)

Qy      96 LysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAsp 115
           |||   |||   ||||||      ||||||   :::      |||:::   :::   :::
Db    100 AAAGGTGGTCAACTGATTCAATCTTTGATAAATCAACATGAGAATCATGTTGATATTAAC 159

Qy     116 AspPheGlnPheVal-------------------------------------TrpAsn 122
           ::::::   |||                                        |||
Db    160 CACTATCGAATTGTTAAAGATGATATAGAAGCTATTACACACCAACTTAAGCTTTGGTTA 219

Qy     123 IleTyrAlaAsnAsn---AspValValValProThrGlyGlyCysAspValSerAlaArg 141
           :::|||      ||||||::::::   |||||||||      :::|||   |||
```

```
Db    220 ACTTCGTCAAATCAACTAGATGTCATTATTACAACGGGAGGAACAGGTATCTCACAAAGA 279

Qy    142 AspValThrVal 145
          ||||||||||:::
Db    280 GATGTGACAATA 291

RESULT 5
US-08-229-781-54
; Sequence 54, Application US/08229781
; Patent No. 5589174
;  GENERAL INFORMATION:
;    APPLICANT:  Yoshinobu OKUNO et al.
;    TITLE OF INVENTION:  ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/229,781
;      FILING DATE:  April 19, 1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/054,016
;      FILING DATE:  April 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1777 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to genomic RNA
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  A2/Aichi/2/68
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:
;      MAP POSITION:
;      UNITS:
;    FEATURE:
;      NAME/KEY:
;      LOCATION:
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:

;    PUBLICATION INFORMATION:
;      AUTHORS:
;      TITLE:
;      JOURNAL:
;      VOLUME:
;      ISSUE:
;      PAGES:
;      DATE:
;      DOCUMENT NUMBER:
;      FILING DATE:
;      PUBLICATION DATE:
;      RELEVANT RESIDUES IN SEQ ID NO:
US-08-229-781-54

Alignment Scores:
Pred. No.:                4.47          Length:          1777
Score:                    77.50         Matches:         44
Percent Similarity:       38.01%        Conservative:    21
Best Local Similarity:    25.73%        Mismatches:      67
Query Match:              9.14%         Indels:          39
DB:                       1             Gaps:            9

US-09-900-575-29_COPY_26_186 (1-161) x US-08-229-781-54 (1-1777)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeu---------SerThrGlnIle 17
          |||      :::|||   ||||||         ||||||             ::::::
Db    304 CCTCATTGTGATGTTTTTCAAAATGAGACATGGGACCTTTTCGTTGAACGCAGCAAAGCT 363

Qy     18 PheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySer 37
          |||       |||   ||||||      :::   ||||||   :::|||:::
Db    364 TTC---AGCAACTGTTACCCTTATGATGTGCCAGATTATGCCTCCCTTAGGTCACTAGTT 420

Qy     38 AlaTyrGlyGlyValLeuSer----------------AsnPheSerGlyThrValLys 51
          |||       |||   |||                  :::::|||      :::
Db    421 GCCTCGTCAGGCACTCTGGAGTTTATCACTGAGGGTTTCACTTGGACTGGGGTCACTCAG 480

Qy     52 TyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsn 71
            ||||||:::                                |||         :::
Db    481 AATGGGGGAAGCAAT-----------GCTTGCAAAAGGGGACCTGGTAGCGGTTTTTTC 528

Qy     72 SerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGly 91
          ||||||   :::      |||
Db    529 AGTAGACTGAAC------TGG---------------------------------------- 543

Qy     92 GlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
            :::   |||:::||||||         ||||||   :::        ||||||
Db    544 ------TTGACCAAATCAGGAAGCACATATCCAGTGCTGAACGTGACTATGCCAAACAAT 597

Qy    112 TyrAsnSerAspAspPheGlnPheValTrpAsnIleTyr---------AlaAsnAsnAsp 128
            |||   |||         ::::::|||   |||:::            |||   :::
Db    598 GACAATTTTGACAAACTA---TACATTTGGGGGATTCACCACCCGAGCACGAACCAAGAA 654

Qy    129 -----------ValValValProThrGlyGlyCysAspValSerAlaArgAspValThr 144
                      |||      :::|||         ||||||   |||
Db    655 CAAACCAGCCTGTATGTTCAAGCATCAGGGAGAGTCACAGTCTCTACCAGGAGAAGCCAG 714

Qy    145 ValThrLeuProAspTyrArgGlySerValPro 155
            |||:::           ||||||   |||
Db    715 CAAACTATAATCCCGAATATCGGGTCCAGACCC 747

RESULT 6
US-08-630-918-54
; Sequence 54, Application US/08630918
; Patent No. 5631350
;  GENERAL INFORMATION:
;    APPLICANT:  Yoshinobu OKUNO et al.
;    TITLE OF INVENTION:  ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack
;      STREET:  805 Fifteenth Street, N.W., #700
;      CITY:  Washington
;      STATE:  D.C.
```

```
;      COUNTRY:  U.S.A.
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/630,918
;      FILING DATE:  April 5, 1996
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/229,781
;      FILING DATE:  April 19, 1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/054,016
;      FILING DATE:  April 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-8850
;      TELEFAX:
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1777 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to genomic RNA
;    ORIGINAL SOURCE:
;      ORGANISM:  A2/Aichi/2/68
US-08-630-918-54

Alignment Scores:
Pred. No.:                4.47          Length:          1777
Score:                    77.50         Matches:         44
Percent Similarity:       38.01%        Conservative:    21
Best Local Similarity:    25.73%        Mismatches:      67
Query Match:              9.14%         Indels:          39
DB:                       1             Gaps:            9

US-09-900-575-29_COPY_26_186 (1-161) x US-08-630-918-54 (1-1777)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeu---------SerThrGlnIle 17
             |||      :::|||   ||||||          ||||||             ::::::
Db     304 CCTCATTGTGATGTTTTTCAAAATGAGACATGGGACCTTTTCGTTGAACGCAGCAAAGCT 363

Qy      18 PheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySer 37
             |||      |||  ||||||      :::   ||||||   :::|||:::
Db     364 TTC---AGCAACTGTTACCCTTATGATGTGCCAGATTATGCCTCCCTTAGGTCACTAGTT 420

Qy      38 AlaTyrGlyGlyValLeuSer----------------AsnPheSerGlyThrValLys 51
             |||      |||   |||                        ::::::|||      :::
Db     421 GCCTCGTCAGGCACTCTGGAGTTTATCACTGAGGGTTTCACTTGGACTGGGGTCACTCAG 480

Qy      52 TyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsn 71
             ||||||:::                          |||      :::
Db     481 AATGGGGGAAGCAAT-----------GCTTGCAAAAGGGGACCTGGTAGCGGTTTTTTC 528

Qy      72 SerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGly 91
             ||||||   ::    |||
Db     529 AGTAGACTGAAC------TGG---------------------------------------- 543

Qy      92 GlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
               :::   |||:::||||||         ||||||   :::         ||||||
Db     544 ------TTGACCAAATCAGGAAGCACATATCCAGTGCTGAACGTGACTATGCCAAACAAT 597

Qy     112 TyrAsnSerAspAspPheGlnPheValTrpAsnIleTyr---------AlaAsnAsnAsp 128
             |||   |||         ::::::|||   |||:::            |||   :::
Db     598 GACAATTTTGACAAACTA---TACATTTGGGGGATTCACCACCCGAGCACGAACCAAGAA 654

Qy     129 ------------ValValValProThrGlyGlyCysAspValSerAlaArgAspValThr 144
                        |||      :::|||         ||||||   |||
Db     655 CAAACCAGCCTGTATGTTCAAGCATCAGGGAGAGTCACAGTCTCTACCAGGAGAAGCCAG 714

Qy     145 ValThrLeuProAspTyrArgGlySerValPro 155
             |||:::          ||||||   |||
Db     715 CAAACTATAATCCCGAATATCGGGTCCAGACCC 747

RESULT 7
US-09-004-422-54
; Sequence 54, Application US/09004422
; Patent No. 6337070
;  GENERAL INFORMATION:
;    APPLICANT:  Yoshinobu OKUNO et al.
;    TITLE OF INVENTION:  POLYPEPTIDES FOR USE IN GENERATING
;    TITLE OF INVENTION:  ANTI-HUMAN INFLUENZA VIRUS ANTIBODIES (AS AMENDED)
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wenderoth, Lind & Ponack, L.L.P.
;      STREET:  2033 K Street, N.W., #800
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20006
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 mb
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/004,422
;      FILING DATE:  January 8, 1998
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/443,862
;      FILING DATE:  May 22, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/229,781
;      FILING DATE:  April 19, 1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/054,016
;      FILING DATE:  April 29, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warren M. Cheek, Jr.
;      REGISTRATION NUMBER:  33,367
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-721-8200
;      TELEFAX:  202-721-8250
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1777 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to genomic RNA
;    HYPOTHETICAL:
;    ANTI-SENSE:
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
;      ORGANISM:  A2/Aichi/2/68
;      STRAIN:
;      INDIVIDUAL ISOLATE:
;      DEVELOPMENTAL STAGE:
;      HAPLOTYPE:
;      TISSUE TYPE:
;      CELL TYPE:
;      CELL LINE:
;      ORGANELLE:
;    IMMEDIATE SOURCE:
;      LIBRARY:
```

```
;       CLONE:
;     POSITION IN GENOME:
;       CHROMOSOME/SEGMENT:
;       MAP POSITION:
;       UNITS:
;     FEATURE:
;       NAME/KEY:
;       LOCATION:
;       IDENTIFICATION METHOD:
;       OTHER INFORMATION:
;     PUBLICATION INFORMATION:
;       AUTHORS:
;       TITLE:
;       JOURNAL:
;       VOLUME:
;       ISSUE:
;       PAGES:
;       DATE:
;       DOCUMENT NUMBER:
;       FILING DATE:
;       PUBLICATION DATE:
;       RELEVANT RESIDUES IN SEQ ID NO:
US-09-004-422-54

Alignment Scores:
Pred. No.:                4.47          Length:          1777
Score:                    77.50         Matches:         44
Percent Similarity:       38.01%        Conservative:    21
Best Local Similarity:    25.73%        Mismatches:      67
Query Match:              9.14%         Indels:          39
DB:                       4             Gaps:            9

US-09-900-575-29_COPY_26_186 (1-161) x US-09-004-422-54 (1-1777)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeu---------SerThrGlnIle 17
           |||      :::|||   ||||||         ||||||             ::::::
Db     304 CCTCATTGTGATGTTTTTCAAAATGAGACATGGGACCTTTTCGTTGAACGCAGCAAAGCT 363

Qy      18 PheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySer 37
           |||      |||   ||||||     :::   ||||||   :::|||:::
Db     364 TTC---AGCAACTGTTACCCTTATGATGTGCCAGATTATGCCTCCCTTAGGTCACTAGTT 420

Qy      38 AlaTyrGlyGlyValLeuSer----------------AsnPheSerGlyThrValLys 51
           |||      |||   |||                  :::::|||      :::
Db     421 GCCTCGTCAGGCACTCTGGAGTTTATCACTGAGGGTTTCACTTGGACTGGGGTCACTCAG 480

Qy      52 TyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsn 71
             |||||||:::                          |||         :::
Db     481 AATGGGGGAAGCAAT-----------GCTTGCAAAAGGGGACCTGGTAGCGGTTTTTTC 528

Qy      72 SerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGly 91
           ||||||   :::      |||
Db     529 AGTAGACTGAAC------TGG------------------------------------ 543

Qy      92 GlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
             :::   |||:::||||||         ||||||   :::          ||||||
Db     544 ------TTGACCAAATCAGGAAGCACATATCCAGTGCTGAACGTGACTATGCCAAACAAT 597

Qy     112 TyrAsnSerAspAspPheGlnPheValTrpAsnIleTyr---------AlaAsnAsnAsp 128
           |||   |||         :::::::|||   |||:::            |||   :::
Db     598 GACAATTTTGACAAACTA---TACATTTGGGGGATTCACCACCCGAGCACGAACCAAGAA 654

Qy     129 -----------ValValValProThrGlyGlyCysAspValSerAlaArgAspValThr 144
                      |||      :::|||          ||||||   |||
Db     655 CAAACCAGCCTGTATGTTCAAGCATCAGGGAGAGTCACAGTCTCTACCAGGAGAAGCCAG 714

Qy     145 ValThrLeuProAspTyrArgGlySerValPro 155
           |||:::          ||||||   |||
Db     715 CAAACTATAATCCCGAATATCGGGTCCAGACCC 747

RESULT 8
US-08-309-341-3/c
; Sequence 3, Application US/08309341
; Patent No. 5594119
;  GENERAL INFORMATION:
;    APPLICANT:  Yaver, Debbie Sue
;    APPLICANT:  Thompson, Sheryl Ann
;    TITLE OF INVENTION:  GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5594119o No. 5594119disk of No. 5594119th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/309,341
;      FILING DATE:  16-SEP-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney, Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  4247.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2002 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  349..411
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join (348..412)
US-08-309-341-3

Alignment Scores:
Pred. No.:                5.33          Length:          2002
Score:                    77.50         Matches:         41
Percent Similarity:       37.75%        Conservative:    16
Best Local Similarity:    27.15%        Mismatches:      52
Query Match:              9.14%         Indels:          42
DB:                       1             Gaps:            6

US-09-900-575-29_COPY_26_186 (1-161) x US-08-309-341-3 (1-2002)

Qy      45 AsnPheSerGlyThrValLysTyrSerGlySerSer---------TyrProPheProThr 61
           |||:::|||   |||         :::||||||              |||:::|||
Db    1220 AACTATAGCAAGACTCAATCATAGACTGGCAGCGAGGAAGGGCGTTGTCCATGGCCTGGC 1161

Qy      62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrp-ProValAl 81
             |||      ||||||   :::   |||   |||      :::|||||    |||
Db    1160 AGGAGCCCTCGTCCAAGACAGCTGGGTAACCACCGTCACCACAGGCCATGGGACGGTAGT 1101

Qy      81 aLeuTyrLeuThr--------------ProValSerSerAlaGlyGlyLeuValIleLy 96
           |||   |||:::                |||::::::      ||||||   :::
Db    1100 ACTCGTACTGAGTGAGACCGTCGGTGAGACCGTTGCCAATAAGAACGGACTGCAGGTTGA 1041

Qy      96 sAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAs 116
                              |||:::::::||||||      |||      |||   ||
Db    1040 T--------------GTTGCGCTTCTTGTGAGACAA------AATCTCCGAAGCAAAGA 1002

Qy     116 pPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGly----- 134
```

```
Db   1001 CGGGGATATAGTGACCAGCATAGGATTCACCGGCAATGTGGAAGTCCTGCTTGGCATACT 942

Qy    135 ------------------------------GlyCysAspValSerAla---------- 140

Db    941 CGGGGAATTGTTTGAAGAAGAGGGTAAGCAAGGCATAGACGTCCTTGCCAGCAGCAACGG 882

Qy    141 ------------------------------ArgAspValThrValThrLeuProAspTy 150

Db    881 TGTCGCTGACAGCAGAGTTGCTGTAAGAGTAACCGACGTTGACAGGCTGGTCAAGGAAGA 822

Qy    150 rArgGlySerValProIleProLeuThrVal 160

Db    821 TCACGGACGCGTTGGAGTTCCAAGCGTAGTC 791

RESULT 9
US-08-608-267-3/c
; Sequence 3, Application US/08608267
; Patent No. 5688663
;  GENERAL INFORMATION:
;    APPLICANT:  Yaver, Debbie Sue
;    APPLICANT:  Thompson, Sheryl Ann
;    TITLE OF INVENTION:  GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5688663o No. 5688663disk of No. 5688663th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/608,267
;      FILING DATE:  28-FEB-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/309,341
;      FILING DATE:  20-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney, Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  4247.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2002 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  349..411
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join (348..412)
US-08-608-267-3

Alignment Scores:
Pred. No.:                5.33          Length:          2002
Score:                    77.50         Matches:         41
Percent Similarity:       37.75%        Conservative:    16
Best Local Similarity:    27.15%        Mismatches:      52
Query Match:              9.14%         Indels:          42

DB:                  1          Gaps:          6

US-09-900-575-29_COPY_26_186 (1-161) x US-08-608-267-3 (1-2002)

Qy      45 AsnPheSerGlyThrValLysTyrSerGlySerSer---------TyrProPheProThr 61

Db    1220 AACTATAGCAAGACTCAATCATAGACTGGCAGCGAGGAAGGGCGTTGTCCATGGCCTGGC 1161

Qy      62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrp-ProValAl 81

Db    1160 AGGAGCCCTCGTCCAAGACAGCTGGGTAACCACCGTCACCACAGGCCATGGGACGGTAGT 1101

Qy      81 aLeuTyrLeuThr---------------ProValSerSerAlaGlyGlyLeuValIleLy 96

Db    1100 ACTCGTACTGAGTGAGACCGTCGGTGAGACCGTTGCCAATAAGAACGGACTGCAGGTTGA 1041

Qy      96 sAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAs 116

Db    1040 T---------------GTTGCGCTTCTTGTGAGACAA------AATCTCCGAAGCAAAGA 1002

Qy     116 pPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGly----- 134

Db    1001 CGGGGATATAGTGACCAGCATAGGATTCACCGGCAATGTGGAAGTCCTGCTTGGCATACT 942

Qy     135 ------------------------------GlyCysAspValSerAla---------- 140

Db     941 CGGGGAATTGTTTGAAGAAGAGGGTAAGCAAGGCATAGACGTCCTTGCCAGCAGCAACGG 882

Qy     141 ------------------------------ArgAspValThrValThrLeuProAspTy 150

Db     881 TGTCGCTGACAGCAGAGTTGCTGTAAGAGTAACCGACGTTGACAGGCTGGTCAAGGAAGA 822

Qy     150 rArgGlySerValProIleProLeuThrVal 160

Db     821 TCACGGACGCGTTGGAGTTCCAAGCGTAGTC 791

RESULT 10
US-08-608-452-3/c
; Sequence 3, Application US/08608452
; Patent No. 5693510
;  GENERAL INFORMATION:
;    APPLICANT:  Yaver, Debbie Sue
;    APPLICANT:  Thompson, Sheryl Ann
;    TITLE OF INVENTION:  GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5693510o No. 5693510disk of No. 5693510th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/608,452
;      FILING DATE:  28-FEB-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/309,341
;      FILING DATE:  20-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney, Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  4247.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  2002 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  349..411
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join (348..412)
US-08-608-452-3

Alignment Scores:
Pred. No.:                5.33          Length:          2002
Score:                    77.50         Matches:         41
Percent Similarity:       37.75%        Conservative:    16
Best Local Similarity:    27.15%        Mismatches:      52
Query Match:              9.14%         Indels:          42
DB:                       1             Gaps:            6

US-09-900-575-29_COPY_26_186 (1-161) x US-08-608-452-3 (1-2002)

Qy      45 AsnPheSerGlyThrValLysTyrSerGlySerSer---------TyrProPheProThr 61
           |||:::|||   |||        :::||||||              |||:::|||
Db    1220 AACTATAGCAAGACTCAATCATAGACTGGCAGCGAGGAAGGGCGTTGTCCATGGCCTGGC 1161

Qy      62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrp-ProValAl 81
             |||      ||||||   :::   |||   |||       :::|||||||    |||
Db    1160 AGGAGCCCTCGTCCAAGACAGCTGGGTAACCACCGTCACCACAGGCCATGGGACGGTAGT 1101

Qy      81 aLeuTyrLeuThr--------------ProValSerSerAlaGlyGlyLeuValIleLy 96
             |||   |||:::                |||::::::      ||||||   :::
Db    1100 ACTCGTACTGAGTGAGACCGTCGGTGAGACCGTTGCCAATAAGAACGGACTGCAGGTTGA 1041

Qy      96 sAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAs 116
                          |||::::::|||||      |||      |||   ||
Db    1040 T--------------GTTGCGCTTCTTGTGAGACAA------AATCTCCGAAGCAAAGA 1002

Qy     116 pPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGly----- 134
           |         |||   :::|||           :::|||   |||      |||
Db    1001 CGGGGATATAGTGACCAGCATAGGATTCACCGGCAATGTGGAAGTCCTGCTTGGCATACT 942

Qy     135 ------------------------------GlyCysAspValSerAla---------- 140
                                           |||   ||||||   |||
Db     941 CGGGGAATTGTTTGAAGAAGAGGGTAAGCAAGGCATAGACGTCCTTGCCAGCAGCAACGG 882

Qy     141 ------------------------------ArgAspValThrValThrLeuProAspTy 150
                                           ||||||         :::   :::
Db     881 TGTCGCTGACAGCAGAGTTGCTGTAAGAGTAACCGACGTTGACAGGCTGGTCAAGGAAGA 822

Qy     150 rArgGlySerValProIleProLeuThrVal 160
             |||   |||   :::|||     |||
Db     821 TCACGGACGCGTTGGAGTTCCAAGCGTAGTC 791

RESULT 11
US-08-608-224-3/c
; Sequence 3, Application US/08608224
; Patent No. 5705376
;  GENERAL INFORMATION:
;    APPLICANT:  Yaver, Debbie Sue
;    APPLICANT:  Thompson, Sheryl Ann
;    TITLE OF INVENTION:  GENE ENCODING CARBOXYPEPTIDASE OF ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5705376o No. 5705376disk of No. 5705376th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/608,224
;      FILING DATE:  28-FEB-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/309,341
;      FILING DATE:  20-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney, Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  4247.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2002 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  349..411
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join (348..412)
US-08-608-224-3

Alignment Scores:
Pred. No.:                5.33          Length:          2002
Score:                    77.50         Matches:         41
Percent Similarity:       37.75%        Conservative:    16
Best Local Similarity:    27.15%        Mismatches:      52
Query Match:              9.14%         Indels:          42
DB:                       1             Gaps:            6

US-09-900-575-29_COPY_26_186 (1-161) x US-08-608-224-3 (1-2002)

Qy      45 AsnPheSerGlyThrValLysTyrSerGlySerSer---------TyrProPheProThr 61
           |||:::|||   |||        :::||||||              |||:::|||
Db    1220 AACTATAGCAAGACTCAATCATAGACTGGCAGCGAGGAAGGGCGTTGTCCATGGCCTGGC 1161

Qy      62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrp-ProValAl 81
             |||      ||||||   :::   |||   |||       :::|||||||    |||
Db    1160 AGGAGCCCTCGTCCAAGACAGCTGGGTAACCACCGTCACCACAGGCCATGGGACGGTAGT 1101

Qy      81 aLeuTyrLeuThr--------------ProValSerSerAlaGlyGlyLeuValIleLy 96
             |||   |||:::                |||::::::      ||||||   :::
Db    1100 ACTCGTACTGAGTGAGACCGTCGGTGAGACCGTTGCCAATAAGAACGGACTGCAGGTTGA 1041

Qy      96 sAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAs 116
                          |||::::::|||||      |||      |||   ||
Db    1040 T--------------GTTGCGCTTCTTGTGAGACAA------AATCTCCGAAGCAAAGA 1002

Qy     116 pPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGly----- 134
           |         |||   :::|||           :::|||   |||      |||
Db    1001 CGGGGATATAGTGACCAGCATAGGATTCACCGGCAATGTGGAAGTCCTGCTTGGCATACT 942

Qy     135 ------------------------------GlyCysAspValSerAla---------- 140
                                           |||   ||||||   |||
Db     941 CGGGGAATTGTTTGAAGAAGAGGGTAAGCAAGGCATAGACGTCCTTGCCAGCAGCAACGG 882

Qy     141 ------------------------------ArgAspValThrValThrLeuProAspTy 150
                                           ||||||         :::   :::
Db     881 TGTCGCTGACAGCAGAGTTGCTGTAAGAGTAACCGACGTTGACAGGCTGGTCAAGGAAGA 822
```

```
Qy    150 rArgGlySerValProIleProLeuThrVal 160
Db    821 TCACGGACGCGTTGGAGTTCCAAGCGTAGTC 791

RESULT 12
US-08-967-149-3/c
; Sequence 3, Application US/08967149
; Patent No. 5939305
;  GENERAL INFORMATION:
;    APPLICANT:  Yaver, Debbie Sue
;    APPLICANT:  Thompson, Sheryl Ann
;    TITLE OF INVENTION:  GENE ENCODING CARBOXYPEPTIDASE OF
;    TITLE OF INVENTION:  ASPERGILLUS NIGER
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5939305o No. 5939305disk of No. 5939305th America, Inc.
;      STREET:  405 Lexington Avenue, Suite 6400
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/967,149
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/608,452
;      FILING DATE:  28-FEB-1996
;      APPLICATION NUMBER:  US 08/309,341
;      FILING DATE:  20-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lowney, Karen A.
;      REGISTRATION NUMBER:  31,274
;      REFERENCE/DOCKET NUMBER:  4247.000-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212 867 0123
;      TELEFAX:  212 867 0298
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2002 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus niger
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  349..411
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join (348..412)
US-08-967-149-3

Alignment Scores:
Pred. No.:                5.33          Length:          2002
Score:                    77.50         Matches:         41
Percent Similarity:       37.75%        Conservative:    16
Best Local Similarity:    27.15%        Mismatches:      52
Query Match:              9.14%         Indels:          42
DB:                       2             Gaps:            6

US-09-900-575-29_COPY_26_186 (1-161) x US-08-967-149-3 (1-2002)

Qy     45 AsnPheSerGlyThrValLysTyrSerGlySerSer---------TyrProPheProThr 61
Db   1220 AACTATAGCAAGACTCAATCATAGACTGGCAGCGAGGAAGGGCGTTGTCCATGGCCTGGC 1161

Qy     62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrp-ProValAl 81
Db   1160 AGGAGCCCTCGTCCAAGACAGCTGGGTAACCACCGTCACCACAGGCCATGGGACGGTAGT 1101

Qy     81 aLeuTyrLeuThr----------ProValSerSerAlaGlyGlyLeuValIleLy 96
Db   1100 ACTCGTACTGAGTGAGACCGTCGGTGAGACCGTTGCCAATAAGAACGGACTGCAGGTTGA 1041

Qy     96 sAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAs 116
Db   1040 T----------GTTGCGCTTCTTGTGAGACAA------AATCTCCGAAGCAAAGA 1002

Qy    116 pPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGly----- 134
Db   1001 CGGGGATATAGTGACCAGCATAGGATTCACCGGCAATGTGGAAGTCCTGCTTGGCATACT 942

Qy    135 ----------GlyCysAspValSerAla---------- 140
Db    941 CGGGGAATTGTTTGAAGAAGAGGGTAAGCAAGGCATAGACGTCCTTGCCAGCAGCAACGG 882

Qy    141 ----------ArgAspValThrValThrLeuProAspTy 150
Db    881 TGTCGCTGACAGCAGAGTTGCTGTAAGAGTAACCGACGTTGACAGGCTGGTCAAGGAAGA 822

Qy    150 rArgGlySerValProIleProLeuThrVal 160
Db    821 TCACGGACGCGTTGGAGTTCCAAGCGTAGTC 791

RESULT 13
US-08-714-918-96
; Sequence 96, Application US/08714918
; Patent No. 6037123
;  GENERAL INFORMATION:
;    APPLICANT:  Benton, Bret
;    APPLICANT:  Lee, Ving
;    APPLICANT:  Malouin, Francois
;    APPLICANT:  Martin, Patrick K.
;    APPLICANT:  Schmid, Molly B.
;    APPLICANT:  Sun, Dongxu
;    TITLE OF INVENTION:  STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
;    TITLE OF INVENTION:  TARGET GENES
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,918
;      FILING DATE:  September 13, 1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/009,102
;      FILING DATE:  December 22, 1995
;      APPLICATION NUMBER:  60/003,798
;      FILING DATE:  September 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  222/005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
```

```
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1916 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-714-918-96

Alignment Scores:
Pred. No.:                8.91          Length:          1916
Score:                    75.50         Matches:         28
Percent Similarity:       41.23%        Conservative:    19
Best Local Similarity:    24.56%        Mismatches:      44
Query Match:              8.90%         Indels:          23
DB:                       3             Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-08-714-918-96 (1-1916)

Qy      11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr------ 28
           :::::::::::::|||:::|||         :::         |||   :::|||
Db     498 ATGGAAATGAATACACGTATTCAAGTAGAACATCCTGTAACTGAAATGGTAACAGGAATT 557

Qy      29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGly 48
           |||   |||   ||||||         |||   |||   ||||||
Db     558 GATTTAGTTAAATTACAATTACAAGTTGCTATGGGTGACGTGTTACCGTATAAACAAGAA 617

Qy      49 ThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgVal 68
             :::|||   :::|||   :::      |||         :::|||   |||
Db     618 GATATTAAATTAACAGGACACGCAATTGAATTTAGAATTAATGCTGAAAATCCT------ 671

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
             |||   :::           |||      :::      ||||||   |||
Db     672 ---TACAAGAACTTTATGCCATCACCAGGTAAAATTGAGCAATATCTTGCACCA------ 722

Qy      89 SerAlaGlyGlyLeuValIleLysAla------------------------------- 97
             ||||||      ::::::
Db     723 ------GGTGGATATGGTGTTCGAATAGAGTCAGCATGTTATACTAATTATACGATACCG 776

Qy      98 ---------GlySerLeuIleAlaValLeuIleLeuArgGln 108
                   |||::::::|||   ||||||:::   :::
Db     777 CCATATTATGATTCGATGGTAGCGAAATTAATCATACATGAA 818

RESULT 14
US-09-265-315-96
; Sequence 96, Application US/09265315
; Patent No. 6187541
;  GENERAL INFORMATION:
;    APPLICANT:  Benton, Bret
;    APPLICANT:  Lee, Ving J.
;    APPLICANT:  Malouin, Francois
;    APPLICANT:  Martin, Patrick K.
;    APPLICANT:  Schmid, Molly B.
;    APPLICANT:  Sun, Dongxu
;    TITLE OF INVENTION:  METHODS OF SCREENING FOR COMPOUNDS
;    TITLE OF INVENTION:  ACTIVE ON STAPHYLOCOCCUS AUREUS
;    TITLE OF INVENTION:  TARGET GENES
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/265,315
;      FILING DATE:  March 9, 1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/714,918
;      FILING DATE:  September 13, 1996
;      APPLICATION NUMBER:  60/009,102
;      FILING DATE:  December 22, 1995
;      APPLICATION NUMBER:  60/003,798
;      FILING DATE:  September 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  240/247
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1916 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-265-315-96

Alignment Scores:
Pred. No.:                8.91          Length:          1916
Score:                    75.50         Matches:         28
Percent Similarity:       41.23%        Conservative:    19
Best Local Similarity:    24.56%        Mismatches:      44
Query Match:              8.90%         Indels:          23
DB:                       4             Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-09-265-315-96 (1-1916)

Qy      11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr------ 28
           :::::::::::::|||:::|||         :::         |||   :::|||
Db     498 ATGGAAATGAATACACGTATTCAAGTAGAACATCCTGTAACTGAAATGGTAACAGGAATT 557

Qy      29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGly 48
           |||   |||   ||||||         |||   |||   ||||||
Db     558 GATTTAGTTAAATTACAATTACAAGTTGCTATGGGTGACGTGTTACCGTATAAACAAGAA 617

Qy      49 ThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgVal 68
             :::|||   :::|||   :::      |||         :::|||   |||
Db     618 GATATTAAATTAACAGGACACGCAATTGAATTTAGAATTAATGCTGAAAATCCT------ 671

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
             |||   :::           |||      :::      ||||||   |||
Db     672 ---TACAAGAACTTTATGCCATCACCAGGTAAAATTGAGCAATATCTTGCACCA------ 722

Qy      89 SerAlaGlyGlyLeuValIleLysAla------------------------------- 97
             ||||||      ::::::
Db     723 ------GGTGGATATGGTGTTCGAATAGAGTCAGCATGTTATACTAATTATACGATACCG 776

Qy      98 ---------GlySerLeuIleAlaValLeuIleLeuArgGln 108
                   |||::::::|||   ||||||:::   :::
Db     777 CCATATTATGATTCGATGGTAGCGAAATTAATCATACATGAA 818

RESULT 15
US-09-265-315-96
; Sequence 96, Application US/09265315
; Patent No. 6187541
;  GENERAL INFORMATION:
;    APPLICANT:  Benton, Bret
;    APPLICANT:  Lee, Ving J.
;    APPLICANT:  Malouin, Francois
;    APPLICANT:  Martin, Patrick K.
;    APPLICANT:  Schmid, Molly B.
;    APPLICANT:  Sun, Dongxu
;    TITLE OF INVENTION:  METHODS OF SCREENING FOR COMPOUNDS
;    TITLE OF INVENTION:  ACTIVE ON STAPHYLOCOCCUS AUREUS
```

```
;    TITLE OF INVENTION:  TARGET GENES
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/265,315
;      FILING DATE:  March 9, 1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/714,918
;      FILING DATE:  September 13, 1996
;      APPLICATION NUMBER:  60/009,102
;      FILING DATE:  December 22, 1995
;      APPLICATION NUMBER:  60/003,798
;      FILING DATE:  September 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  240/247
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1916 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-265-315-96

Alignment Scores:
Pred. No.:                8.91          Length:          1916
Score:                    75.50         Matches:         28
Percent Similarity:       41.23%        Conservative:    19
Best Local Similarity:    24.56%        Mismatches:      44
Query Match:              8.90%         Indels:          23
DB:                       4             Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-09-265-315-96 (1-1916)

Qy      11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr------ 28
           :::::::::::::|||:::|||       :::         |||   :::|||
Db     498 ATGGAAATGAATACACGTATTCAAGTAGAACATCCTGTAACTGAAATGGTAACAGGAATT 557

Qy      29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGly 48
           |||   |||   ||||||         |||   |||   ||||||
Db     558 GATTTAGTTAAATTACAATTACAAGTTGCTATGGGTGACGTGTTACCGTATAAACAAGAA 617

Qy      49 ThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgVal 68
           :::|||   :::|||   :::       |||         :::|||   |||
Db     618 GATATTAAATTAACAGGACACGCAATTGAATTTAGAATTAATGCTGAAAATCCT------ 671

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::         |||   :::       ||||||   |||
Db     672 ---TACAAGAACTTTATGCCATCACCAGGTAAAATTGAGCAATATCTTGCACCA------ 722

Qy      89 SerAlaGlyGlyLeuValIleLysAla------------------------------- 97
           ||||||      ::::::
Db     723 ------GGTGGATATGGTGTTCGAATAGAGTCAGCATGTTATACTAATTATACGATACCG 776

Qy      98 ---------GlySerLeuIleAlaValLeuIleLeuArgGln 108
                    |||:::::::|||   ||||||:::  :::
Db     777 CCATATTATGATTCGATGGTAGCGAAATTAATCATACATGAA 818

RESULT 16
US-09-266-417-96
; Sequence 96, Application US/09266417
; Patent No. 6228588
;  GENERAL INFORMATION:
;    APPLICANT:  Benton, Bret
;    APPLICANT:  Lee, Ving J.
;    APPLICANT:  Malouin, Francois
;    APPLICANT:  Martin, Patrick K.
;    APPLICANT:  Schmid, Molly B.
;    APPLICANT:  Sun, Dongxu
;    TITLE OF INVENTION:  METHODS OF SCREENING FOR COMPOUNDS
;    TITLE OF INVENTION:  ACTIVE ON STAPHYLOCOCCUS AUREUS
;    TITLE OF INVENTION:  TARGET GENES
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/266,417
;      FILING DATE:  March 9, 1999
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/714,918
;      FILING DATE:  September 13, 1996
;      APPLICATION NUMBER:  60/009,102
;      FILING DATE:  December 22, 1995
;      APPLICATION NUMBER:  60/003,798
;      FILING DATE:  September 15, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  240/248
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1916 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-266-417-96

Alignment Scores:
Pred. No.:                8.91          Length:          1916
Score:                    75.50         Matches:         28
Percent Similarity:       41.23%        Conservative:    19
Best Local Similarity:    24.56%        Mismatches:      44
Query Match:              8.90%         Indels:          23
DB:                       4             Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-09-266-417-96 (1-1916)

Qy      11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr------ 28
           :::::::::::::|||:::|||       :::         |||   :::|||
Db     498 ATGGAAATGAATACACGTATTCAAGTAGAACATCCTGTAACTGAAATGGTAACAGGAATT 557
```

```
Qy      29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGly 48
           |||   |||   ||||||        |||   |||   ||||||
Db     558 GATTTAGTTAAATTACAATTACAAGTTGCTATGGGTGACGTGTTACCGTATAAACAAGAA 617

Qy      49 ThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgVal 68
           :::|||   :::|||   :::      |||         :::|||   |||
Db     618 GATATTAAATTAACAGGACACGCAATTGAATTTAGAATTAATGCTGAAAATCCT------ 671

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::            |||      :::      ||||||   |||
Db     672 ---TACAAGAACTTTATGCCATCACCAGGTAAAATTGAGCAATATCTTGCACCA------ 722

Qy      89 SerAlaGlyGlyLeuValIleLysAla------------------------------- 97
           ||||||      ::::::
Db     723 ------GGTGGATATGGTGTTCGAATAGAGTCAGCATGTTATACTAATTATACGATACCG 776

Qy      98 ---------GlySerLeuIleAlaValLeuIleLeuArgGln 108
                   |||:::::::|||   ||||||:::   :::
Db     777 CCATATTATGATTCGATGGTAGCGAAATTAATCATACATGAA 818
```

RESULT 17
US-09-641-638-65
; Sequence 65, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CP1
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 65
; LENGTH: 420
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 364
; OTHER INFORMATION: 12-230-364 : polymorphic base C or T
; NAME/KEY: misc_binding
; LOCATION: 344..363
; OTHER INFORMATION: 12-230-364.mis1, potential
; NAME/KEY: misc_binding
; LOCATION: 365..384
; OTHER INFORMATION: 12-230-364.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 1..20
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 401..420
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 352..376
; OTHER INFORMATION: 12-230-364 potential probe
US-09-641-638-65

Alignment Scores:

Pred. No.:	1.1	Length:	420
Score:	75.00	Matches:	31
Percent Similarity:	40.16%	Conservative:	20
Best Local Similarity:	24.41%	Mismatches:	58
Query Match:	8.84%	Indels:	18
DB:	4	Gaps:	3

US-09-900-575-29_COPY_26_186 (1-161) x US-09-641-638-65 (1-420)

```
Qy      32 ThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLys 51
           :::::::::||||||      :::   |||:::|||               :::
Db      26 TCTATAAAAAGGGGGCAATGCTTCAAAGGATTGTTGGTGAGACTTCAGTGAAGTAAAGCT 85

Qy      52 TyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsn 71
           :::   |||   |||            ||||||      |||   :::      |||
Db      86 TGGGAAGGACCCAGC-----------ACAACTCCAGGCACAGGGCAGGCGTCCTACAAA 133

Qy      72 SerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGly 91
           |||                              ::::::      ||||||::::::
Db     134 TGTAGGTTCCCATCCATTTCATTCCATCCCATTTTCTAGCCACCGGTAGCCACCCTTTCA 193

Qy      92 GlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
                                                  |||||||||:::|||:::
Db     194 TCT-----------------------------------CTACGCCAGTCCAATTCT 214

Qy     112 TyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValVal 131
           |||                 ||||||                 :::|||
Db     215 ACTCCAAGTCACTATGCCCCTCCTGTCTGGATTTGTGAGGGGCAGTCAGACCAGAAGGGG 274

Qy     132 ProThrGlyGly---CysAspValSerAlaArgAspValThrValThrLeuProAspTyr 150
           |||   |||       |||   :::   |||      |||      ||||||::::::
Db     275 CCAGAAGGTGATATTTGTCATCTCCCTGCTGCTATACCAACCTCTCAGCTACCTGAGTTT 334

Qy     151 ArgGlySerValProIlePro 157
               ||||||:::   :::|||
Db     335 TTTGGGTCCCTTTGTCTGCCT 355
```

RESULT 18
US-08-684-862-11
; Sequence 11, Application US/08684862
; Patent No. 5759541
; GENERAL INFORMATION:
; APPLICANT: Bach, Alfred
; APPLICANT: Hillen, Heinz
; APPLICANT: Bialojan, Siegfried
; TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 5.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,862
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,705
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,040
; FILING DATE: 30-DEC-1992
; APPLICATION NUMBER: PCT/EP91/01361
; FILING DATE: 19-JUL-1991
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 957 base pairs

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Agkistrodon rhodostoma
;     FEATURE:
;       LOCATION:  210 to 911
;       OTHER INFORMATION:  the coding region shown in (2)(ix)(B)
;       OTHER INFORMATION:  codes for the protein of SEQ ID NO: 6
US-08-684-862-11

Alignment Scores:
Pred. No.:                3.7          Length:          957
Score:                    75.00        Matches:         28
Percent Similarity:       42.86%       Conservative:    11
Best Local Similarity:    30.77%       Mismatches:      39
Query Match:              8.84%        Indels:          13
DB:                       1            Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-08-684-862-11 (1-957)

Qy      11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThrAspTyr 30
           ::::::|||                |||||      |||||||||   :::   :::|||
Db     633 ATTAACCTACTCGATGATGCAGAGTGTCATGCAGGTTACCCTGAGGTGCTGGCAGAATAC 692

Qy      31 ValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrVal 50
              ||||||      |||         ||||||                      |||
Db     693 AGAACATTGTGTGCAGGTATCGTGCAAGGAGGC--------------AAAGATACATGT 737

Qy      51 LysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValVal--- 69
             |||      |||      |||             :::             :::|||
Db     738 ATGTATGAC---TCTGGAGGACCTCTCATCTGTAATGAACAAGTCCAGGGCATTGTATCT 794

Qy      70 TyrAsnSerArgThrAspLysProTrpPro-ValAlaLeuTyrLeuThrProValSerSe 89
           |||   :::                |||  ||||||     |||:::   ||||||::
Db     795 TATGGGGCGCAT--------------CCTTGTGGCCAACCTCTTAAGCCTGGTATCTAC 839

Qy      89 rAlaGlyGlyLeuValIleLysAlaGlySer 99
           :   |||   ::::::|||      ||||||
Db     840 ACCAGGCTCCATGATTATAATGACTGGATCA 870

RESULT 19
US-08-169-927-1
; Sequence 1, Application US/08169927
; Patent No. 5783441
;  GENERAL INFORMATION:
;    APPLICANT:  Carl, Mitchell
;    APPLICANT:  Dobson, Michael E.
;    APPLICANT:  Ching, Wei Mei
;    APPLICANT:  Dasch, Gregory A
;    TITLE OF INVENTION:  Gene and Protein Applicable to the
;    TITLE OF INVENTION:  Preparation of Vaccines for Rickettsia prowazekii and
;    TITLE OF INVENTION:  Rickettsia typhi and the Detection of Both
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Counsel, Naval Medical R & D Command
;      STREET:  Bldg. 1, T-12, 8901 Wisconsin Ave.
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20889-5606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/169,927
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/742,128
;      FILING DATE:  08/09/91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, A. David
;      REGISTRATION NUMBER:  24,743
;      REFERENCE/DOCKET NUMBER:  75,976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-1022
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Rickettsia prowazekii
;      STRAIN:  Breinl
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  340..345
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  363..368
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  391..5226
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  379..386
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5270..5306
;    PUBLICATION INFORMATION:
;      AUTHORS:  Carl, M.
;      AUTHORS:  Dobson, M. E.
;      AUTHORS:  Ching, W. M.
;      AUTHORS:  Dasch, G. A.
;      TITLE:  Characterization of the gene encoding the
;      TITLE:  protective S-layer protein of Rickettsia
;      TITLE:  prowazekii; presence of a truncated identical
;      TITLE:  homolog in rickettsia typhi
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      DATE:  1990
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 5319
US-08-169-927-1

Alignment Scores:
Pred. No.:                46.4         Length:          5319
Score:                    75.00        Matches:         31
Percent Similarity:       36.64%       Conservative:    17
Best Local Similarity:    23.66%       Mismatches:      47
Query Match:              8.84%        Indels:          36
DB:                       1            Gaps:            5

US-09-900-575-29_COPY_26_186 (1-161) x US-08-169-927-1 (1-5319)

Qy      31 ValThrLeuGlnArgGlySerAlaTyrGlyGlyVal------LeuSerAsnPheSerGly 48
           :::|||:::          :::   |||   |||:::            |||      |||
Db    3112 ATTACTGTTAATGCTAATACATTATATTCAGGTATCAGAACTACAAAAAATAATCAAGGT 3171

Qy      49 ThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgVal 68
           ||||||      ||||||                                      |||
Db    3172 ACTGTGACACTTAGTGGTGGT----------------ATGCCTAATAATCCTGGTACA 3213

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           :::|||
Db    3214 ATTTAT---------------------------------------------------- 3219

Qy      89 SerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGln 108
```

```
Db   3220 ---------GGTTTAGGTTTAGAGAATGGTAGT------------CCAAAGTTAAAACAA 3258

Qy    109 ---------ThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAla 125

Db   3259 GTGACATTTACTACAGATTATAACAACTTAGGTAGTATTATTGCAAATAATGTAACAATT 3318

Qy    126 AsnAsnAspValValValProThrGlyGlyCysAspValSerAlaArgAspValThrVal 145

Db   3319 AATGATGATGTAACTCTTACTACAGGAGGTATAGCAGGGACAGATTTTGACGCTAAAATT 3378

Qy    146 ThrLeuProAspTyrArgGlySerValProIle 156

Db   3379 ACTCTTGGAAGTGTTAACGGTAACGCTAACGTA 3411
```

RESULT 20
US-09-453-702B-102
; Sequence 102, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7886
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-09-453-702B-102

Alignment Scores:

Pred. No.:	95.8	Length:	7886
Score:	74.50	Matches:	27
Percent Similarity:	42.86%	Conservative:	30
Best Local Similarity:	20.30%	Mismatches:	57
Query Match:	8.79%	Indels:	19
DB:	4	Gaps:	3

US-09-900-575-29_COPY_26_186 (1-161) x US-09-453-702B-102 (1-7886)

```
Qy      3 ValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHisAsnAsp 22

Db   7308 GTCCATTTGGGCAATACGAAAGTCGATGATACGGCGAACCATCTCTTCAATAATGGCGTC 7367

Qy     23 TyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyVal 42

Db   7368 TTTGTTGGCGAAGTAACGATATATCTGACCCACGCTAAGCTGAGC--------------- 7412

Qy     43 LeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThr 62

Db   7413 ---------TTCGCTGGCAATTTGCGACATACTGGCGGCATGAAACCCGCAGCGACGAAA 7463

Qy     63 SerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeu 82

Db   7464 ACACCGACGGGCGGCGTT--------------------AATGATTTCATCCTGACGCTG 7502

Qy     83 TyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAla 102

Db   7503 TTTGTGGCGTATTTCCAGCTTGCTGGTCATTTCTGTCTCCTGGCCTATCACTTGGTGAGA 7562

Qy    103 ValLeuIleLeuArgGlnThrAsn-----------AsnTyrAsnSerAspAspPheGln 118

Db   7563 ATGATCGTTCTCATTTTTGCTGATTATACGTTTCTCCAAAGTAATGGAAAGTATTTTTAT 7622

Qy    119 PheValTrpAsnIleTyrAlaAsnAsnAspValValVal 131

Db   7623 CTTATTCATGGTGCATATTCTGATAATGAAGAGATTGTG 7661
```

RESULT 21
US-09-245-041-3/c
; Sequence 3, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Nagle, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
; FILE REFERENCE: 7853-136
; CURRENT APPLICATION NUMBER: US/09/245,041
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/093,630
; EARLIER FILING DATE: 1998-07-21
; EARLIER APPLICATION NUMBER: 60/104,978
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 17056
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-245-041-3

Alignment Scores:

Pred. No.:	345	Length:	17056
Score:	74.00	Matches:	30
Percent Similarity:	43.56%	Conservative:	14
Best Local Similarity:	29.70%	Mismatches:	38
Query Match:	8.73%	Indels:	19
DB:	4	Gaps:	3

US-09-900-575-29_COPY_26_186 (1-161) x US-09-245-041-3 (1-17056)

```
Qy      2 ValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHisAsn 21

Db   9975 GTGCTCAATCCTCACCAAGGTCTGATCCTTGACCTTTGAGAATCTTTCTTC--------- 9925

Qy     22 AspTyrProGluThrIleThrAspTyr-ValThrLeuGlnArgGlySerAlaTyrGlyGl 41

Db   9924 --------------CTTTCTACATATTTTGACCCTTGAGGAGCTTATAGTCTTGGGAGG 9880

Qy     41 yValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProTh 61

Db   9879 AGAGAGATACACTGATTTGGGGATGACTAGGTATGATGGTTAATCTTATGTAGCTTGATT 9820
```

```
Qy      61 rThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProValAl 81
           ||||||||||||||||           :::       ||||||  :::||
Db    9819 GGATTGAGAGACACCTAGAGTT-----------CAAGATGGTATACCTTGGTATATTGC 9772

Qy      81 aLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeuIl 101
           |              |||   ||||||||||||   :::   :::      :::::
Db    9771 T----------------AGCTAGGCTGGGGGGCTAGATCTGGCATCAAATCAAGTTGT 9730

Qy     101 e 101
           :
Db    9729 A 9729

RESULT 22
US-09-453-702B-259
; Sequence 259, Application US/09453702B
; Patent No. 6365723
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/453,702B
;             FILING DATE: 03-Dec-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 259:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4162 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 259:
US-09-453-702B-259

Alignment Scores:
Pred. No.:                57.6          Length:          4162
Score:                    73.00         Matches:         37
Percent Similarity:       36.73%        Conservative:    17
Best Local Similarity:    25.17%        Mismatches:      65
Query Match:              8.61%         Indels:          28
DB:                       4             Gaps:            5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-453-702B-259 (1-4162)

Qy      22 AspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGly 41
           |||:::        :::        |||:::  |||  ||||||  :::

Db     792 GACTGGGACGGCGCGGTGGATGTGCTGGTGTCCAAACAGTCGGGCAGCGTGTTC------ 845

Qy      42 ValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThr 61
           ||||||      :::   ||||||  |||  |||::|||
Db     846 ---CTGAGCCGCAACACCAACACGGTGAGCTAC---GGCACTTCGCTACACCTGCGCATC 899

Qy      62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProValAla 81
           |||               ||||||        |||  |||
Db     900 ACCGATCCCAACGGCATTAACGTCTATTACGGCAATACCGTGAAG-------------- 944

Qy      82 LeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeuIle 101
                                          |||        ||||||  |||:::
Db     945 --------------------------------CTGTACAACTCGGCGGGAGTGCTGGTC 971

Qy     102 AlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrp 121
           |||     |||:::       :::          :::|||       :::
Db     972 GCCACGCAAATCATCAACCCGCAGTCGGGTATGGGGGTTAACGACACCTCGGCGCTGGTC 1031

Qy     122 AsnIleTyr----------------AlaAsnAsnAspValValValProThrGlyGly 135
           |||  |||                        |||  |||::::::  :::  |||
Db    1032 AACTTCTACGGGCTGAATGCCGGAGAAACCTACAACGCGGTGCTGATCAAATCCACCGGC 1091

Qy     136 CysAspValSer------AlaArgAspValThrValThrLeuProAspTyrArgGlySer 153
                  |||       |||     |||                 ||||||
Db    1092 ACCACCGCCAGCAATATGACCAGACGGTCAACACCAGCTGGGGCGGTTTACAGGCCACCG 1151

Qy     154 ValProIleProLeuThrVal 160
           :::|||:::   ||||||:::
Db    1152 ATGCCACTCACGCTTACGATC 1172

RESULT 23
US-08-908-643C-1
; Sequence 1, Application US/08908643C
; Patent No. 6120995
;   GENERAL INFORMATION:
;        APPLICANT: Waldman, Scott A.
;                   Pearlman, Joshua M.
;                   Barber, Michael T.
;                   Schultz, Stephanie
;                   Parkinson, Scott J.
;        TITLE OF INVENTION: COMPOSITIONS THAT SPECIFICALLY BIND TO
;                            COLORECTAL CANCER CELLS AND METHODS OF
;                            USING THE SAME
;        NUMBER OF SEQUENCES: 85
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6120995ris LLP
;             STREET: One Liberty Place - 46th Floor
;             CITY: Philadelphia
;             STATE: PA
;             COUNTRY: U.S.A.
;             ZIP: 19103
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch disk, 1.44 Mb
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WordPerfect 6.1
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/908,643C
;             FILING DATE: 07-Aug-1997
;             CLASSIFICATION: N/A
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: <Unknown>
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Mark Deluca
;             REGISTRATION NUMBER: 33,229
;             REFERENCE/DOCKET NUMBER: TJU-2209
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 215-568-3100
;             TELEFAX: 215-568-3439
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1636 bases
```

```
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-908-643C-1

Alignment Scores:
Pred. No.:                16.8          Length:          1636
Score:                    72.50         Matches:         39
Percent Similarity:       38.51%        Conservative:    23
Best Local Similarity:    24.22%        Mismatches:      60
Query Match:              8.55%         Indels:          40
DB:                       3             Gaps:            6

US-09-900-575-29_COPY_26_186 (1-161) x US-08-908-643C-1 (1-1636)

Qy      19 CysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAla 38
           ||||||  ||||||  ||||||:::|||     ::::::      |||
Db     340 TGTCATGT-GACTATAAAGAAACCTTAACCAGGCTGATGTCTCCAGCTAGAAAGTTGATA 398

Qy      39 TyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrPro 58
           |||         |||   ||||||             :::      :::      |||
Db     399 TAC------TTCTTGGTTAACTTT--------------TGGAAAACCAACGATCTGCCC 437

Qy      59 PheProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrp 78
           |||   |||   |||   :::          ||||||   :::   |||:::
Db     438 TTCAAAACTTATTCCTGGAGCACTTCGTATGTTTACAAGAATGGTACAGAAACTGAG--- 494

Qy      79 ProValAlaLeuTyrLeuThrProValSerSerAlaGly-GlyLeuValIleLysAlaGl 98
                 |||:::|||   |||      |||   |||    :::          :::
Db     495 ------GGACTGTTTCTGGTACCTTAATGCTCTGGAGGCTAGCGTTTCCTATTTCTCCCA 548

Qy      98 ySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGl 118
              |||         ::::::|||||||||          :::
Db     549 CGAACTCGGCTTTAAGGTGGTGTTAAGACAAGATAAGGAGTTTCAG-----------GA 596

Qy     118 nPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAsp-- 137
           :::   :::          :::::|||::::::      ||||||   :::
Db     597 TATCTTAATGGACCACAACAGGAAAAGCAATGTGATTATTATGTGTGGTGGTCCAGAGTT 656

Qy     138 ------------------------ValSerAlaArgAspVal---------------- 143
                                    |||   |||:::
Db     657 CCTCTACAAGCTGAAGGGTGACCGAGCAGTGGCTGAAGACATTGTCATTATTCTAGTGGA 716

Qy     144 ---------------------------ThrValThrLeuProAspTyrArgGlySerVa 154
                                       ||||||   |||||||||      :::||
Db     717 TCTTTTCAATGACCAGTACTTGGAGGACAATGTCACAGCCCCTGACTATATGAAAAATGT 776

Qy     154 l 154
           |
Db     777 C 777

RESULT 24
US-08-714-402-1
; Sequence 1, Application US/08714402
; Patent No. 5910441
;  GENERAL INFORMATION:
;    APPLICANT:  ROCHA, Claudia
;    APPLICANT:  FISCHETTI, Vincent A.
;    TITLE OF INVENTION:  FIBRONECTIN AND FIBRINOGEN BINDING
;    TITLE OF INVENTION:  PROTEIN FROM GROUP A STREPTOCOCCI
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BURNS, DOANE, SWECKER & MATHIS
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,402
;      FILING DATE:  16-SEP-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McGowan, Malcolm K.
;      REGISTRATION NUMBER:  39,300
;      REFERENCE/DOCKET NUMBER:  016921-097
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3531 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-714-402-1

Alignment Scores:
Pred. No.:                60.4          Length:          3531
Score:                    72.00         Matches:         29
Percent Similarity:       37.21%        Conservative:    19
Best Local Similarity:    22.48%        Mismatches:      31
Query Match:              8.49%         Indels:          50
DB:                       2             Gaps:            7

US-09-900-575-29_COPY_26_186 (1-161) x US-08-714-402-1 (1-3531)

Qy      26 ThrIleThrAspTyrVal---------------------------------------- 31
           |||   |||||||||:::
Db     940 ACTTTCACAGATTATATTGCGGGTTTAGATAAAGTCCAGTTGTCTGCAGAATTGAGCTTA 999

Qy      32 -----------------ThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsn 45
                            |||:::      |||              :::||||||
Db    1000 TTCCTAGAGAATAAGGAAGTGTTGGAAAATACTAGT--------------ATCTCAAAT 1044

Qy      46 PheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThr 65
           |||     |||:::              :::|||                     |||
Db    1045 TTTAAGAGTACCATAGGTGGGCAGGAGATCACCTAT--------------AAAGGAACG 1089

Qy      66 ProArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThr 85
              |||:::|||   :::   :::   |||           :::   |||:::|||
Db    1090 GTTAATGTTCTTTATGGAAATGAGAGCACTAAAGAA---------AGCAATTATATTACT 1140

Qy      86 ---ProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeu 104
              :::|||:::   ||||||   :::
Db    1141 AATGGATTGAGCAATGTGGGTGGGAGTATT---------------------------- 1170

Qy     105 IleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyr 124
                          :::||||||::::::      :::|||||||||   :::|||
Db    1171 --------------GAAAGCTACAACACCGAAACGGGAGAATTTGTCTGGTATGTTTAT 1215

Qy     125 AlaAsn---AsnAspValValValPro 132
           |||   |||         :::|||
Db    1216 GTCAATCCAAACCGTACCAATATTCCT 1242

RESULT 25
US-09-327-536-1
; Sequence 1, Application US/09327536
; Patent No. 6355477
; GENERAL INFORMATION:
;  APPLICANT:  FISCHETTI, Vincent A.
;  APPLICANT:  ROCHA, Claudia
;  TITLE OF INVENTION:  FIBRONECTIN AND FIBRINOGEN BINDING PROTEIN FROM GROUP A
;  TITLE OF INVENTION:  STREPTOCOCCI
;  FILE REFERENCE:  022927-008
;  CURRENT APPLICATION NUMBER: US/09/327,536
;  CURRENT FILING DATE:  1999-06-08
;  PRIOR APPLICATION NUMBER:  US 08/714,402
;  PRIOR FILING DATE:  1996-09-16
```

```
;  NUMBER OF SEQ ID NOS:  2
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3698
;   TYPE: DNA
;   ORGANISM: SFFBP gene
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (196)..(3681)
US-09-327-536-1

Alignment Scores:
Pred. No.:                64.7          Length:          3698
Score:                    72.00         Matches:         29
Percent Similarity:       37.21%        Conservative:    19
Best Local Similarity:    22.48%        Mismatches:      31
Query Match:              8.49%         Indels:          50
DB:                       4             Gaps:            7

US-09-900-575-29_COPY_26_186 (1-161) x US-09-327-536-1 (1-3698)

Qy      26 ThrIleThrAspTyrVal---------------------------------------- 31
           |||   |||||||||:::
Db     940 ACTTTCACAGATTATATTGCGGGTTTAGATAAAGTCCAGTTGTCTGCAGAATTGAGCTTA 999

Qy      32 -----------------ThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsn 45
                            |||:::      |||               :::||||||
Db    1000 TTCCTAGAGAATAAGGAAGTGTTGGAAAATACTAGT--------------ATCTCAAAT 1044

Qy      46 PheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThr 65
           |||      |||:::                 :::|||                    |||
Db    1045 TTTAAGAGTACCATAGGTGGGCAGGAGATCACCTAT--------------AAAGGAACG 1089

Qy      66 ProArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThr 85
             |||:::|||   :::   :::   |||           :::   |||:::|||
Db    1090 GTTAATGTTCTTTATGGAAATGAGAGCACTAAAGAA---------AGCAATTATATTACT 1140

Qy      86 ---ProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeu 104
              :::|||:::   ||||||  :::
Db    1141 AATGGATTGAGCAATGTGGGTGGGAGTATT---------------------------- 1170

Qy     105 IleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyr 124
                       :::|||||::::::      :::||||||||||   :::|||
Db    1171 --------------GAAAGCTACAACACCGAAACGGGAGAATTTGTCTGGTATGTTTAT 1215

Qy     125 AlaAsn---AsnAspValValValPro 132
           |||   |||         :::|||
Db    1216 GTCAATCCAAACCGTACCAATATTCCT 1242

RESULT 26
US-09-453-702B-39
; Sequence 39, Application US/09453702B
; Patent No. 6365723
;   GENERAL INFORMATION:
;        APPLICANT: Blattner, Frederick R.
;                   Burland,     Valerie
;                   Perna,  Nicole T.
;                   Plunkett, Guy
;                   Welch,  Rod
;        TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
;        NUMBER OF SEQUENCES: 265
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Quarles & Brady
;             STREET: 1 South Pinckney Street
;             CITY: Madison
;             STATE: WI
;             COUNTRY: US
;             ZIP: 53701-2113
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: Word Perfect 8.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/453,702B
;             FILING DATE: 03-Dec-1999
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/110,955
;             FILING DATE: 04-DEC-1998
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Seay, Nicholas J.
;             REGISTRATION NUMBER: 27386
;             REFERENCE/DOCKET NUMBER: 960296.95017
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (608) 251-5000
;             TELEFAX: (608) 251-9166
;   INFORMATION FOR SEQ ID NO: 39:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 25165
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA  (genomic)
;        SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-453-702B-39

Alignment Scores:
Pred. No.:                1.09e+03      Length:          25165
Score:                    72.00         Matches:         39
Percent Similarity:       37.84%        Conservative:    17
Best Local Similarity:    26.35%        Mismatches:      63
Query Match:              8.49%         Indels:          29
DB:                       4             Gaps:            5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-453-702B-39 (1-25165)

Qy      22 AspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGly 41
           |||:::          :::          |||:::   |||   ||||||   :::
Db  20797 GACTGGGACGGCGCGGTGGATGTGCTGGTGTCCAAACAGTCGGGCAGCGTGTTC------ 20850

Qy      42 ValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThr 61
              ||||||      :::   ||||||   |||   |||:::|||
Db  20851 ---CTGAGCCGCAACACCAACACGGTGAGCTAC---GGCACTTCGCTACACCTGCGCATC 20904

Qy      62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProValAla 81
           |||                  ||||||          |||   |||
Db  20905 ACCGATCCCAACGGCATTAACGTCTATTACGGCAATACCGTGAAG-------------- 20949

Qy      82 LeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeuIle 101
                                        |||           ||||||   |||:::
Db  20950 ------------------------------CTGTACAACTCGGCGGGAGTGCTGGTC 20976

Qy     102 AlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrp 121
           |||      |||:::        :::          :::|||         :::
Db  20977 GCCACGCAAATCATCAACCCGCAGTCGGGTATGGGGGTTAACGACACCTCGGCGCTGGTC 21036

Qy     122 AsnIleTyr-----------------AlaAsnAsnAspValValValProThrGlyGly 135
           |||   |||                  |||   |||:::::::   :::   |||
Db  21037 AACTTCTACGGGCTGAATGCCGGAGAAACCTACAACGCGGTGCTGATCAAATCCACCGGC 21096

Qy     136 CysAspValSerAlaArgAspValThrVal-ThrLeuProAsp------TyrArgGlySe 153
                 |||      |||   |||||| |||                 ||||||
Db  21097 ACCACCGCCAGCAATATCGACCAGACGGTCAACACCAGCTGGGGCGGTTTACAGGCCACC 21156

Qy     153 rValProIleProLeuThrVal 160
           :::|||:::   ||||||:::
Db  21157 GATGCCACTCACGCTTACGATC 21178

RESULT 27
US-07-918-023-1
; Sequence 1, Application US/07918023
; Patent No. 5427934
;  GENERAL INFORMATION:
;    APPLICANT:  Zimmermann, Thomas
;    APPLICANT:  Robins, Karen
```

```
;    APPLICANT:  Birch, Olwen
;    APPLICANT:  Bohlen, Elisabeth
;    TITLE OF INVENTION:  GENETIC ENGINEERING PROCESS FOR THE
;    TITLE OF INVENTION:  PRODUCTION OF S-(+)-2,2-DIMETHYLCYCLOPROPANECARBOXAMIDE
;    TITLE OF INVENTION:  BY MICROORGANISMS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fisher, Christen & Sabol
;      STREET:  2000 M Street, N.W., Suite 590
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/918,023
;      FILING DATE:  19920724
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2247/91
;      FILING DATE:  26-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Marsh, Virgil H.
;      REGISTRATION NUMBER:  23,083
;      REFERENCE/DOCKET NUMBER:  LP 1521
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 659-2000
;      TELEFAX:  (202) 659-2015
;      TELEX:  248748
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1843 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Comamonas acidovorans
;      STRAIN:  A:18
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  289..1566
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  IE 9224406
;      FILING DATE:  23-JUL-1992
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 1843
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  JP 198717/92
;      FILING DATE:  24-JUL-1992
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 1843
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  PL P 295408
;      FILING DATE:  24-JUL-1992
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 1843
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  RO 92-01033
;      FILING DATE:  24-JUL-1992
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 1843
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  SU UNKNOWN
;      FILING DATE:  24-JUL-1992
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 1843
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  CS PV2323-92
;      FILING DATE:  24-JUL-1992
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 1843
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  HU P9202439
;      FILING DATE:  24-JUL-1992
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 1843
US-07-918-023-1

Alignment Scores:
Pred. No.:                26.8          Length:          1843
Score:                    71.50         Matches:         22
Percent Similarity:       42.67%        Conservative:    10
Best Local Similarity:    29.33%        Mismatches:      40
Query Match:              8.43%         Indels:          3
DB:                       1             Gaps:            1

US-09-900-575-29_COPY_26_186 (1-161) x US-07-918-023-1 (1-1843)

Qy      24 ProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeu 43
           |||   |||                |||||||||||||   :::   :::|||
Db    1421 CCCGCCACGGCCTTTGCGGCCCCCACGCTGCAACGCATGGCGCATTTCGGCTCCGATGCC 1480

Qy      44 SerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSer 63
           |||||||||   ::::::|||         :::   ||||||      |||
Db    1481 GAGCTGTTCTCGGGCATGCTGCGCTAC---------ACCTGCCCGTTCGACCTCACGGGC 1531

Qy      64 GluThrProArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyr 83
                        :::         ||||||   :::      |||||||||
Db    1532 AGCCCCACGATCACGCTGCCCGGCGGACGCACTTCTGAGGGCGCGCCCGTGGCCTTCCAG 1591

Qy      84 LeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGly 98
                                   |||::::::::::||||||
Db    1592 TTCGTGGCCCCCGACTTCCGCGAAGACCTGCTGGTGCGCGCGGGC 1636

RESULT 28
US-08-542-003-1/c
; Sequence 1, Application US/08542003
; Patent No. 5864013
;  GENERAL INFORMATION:
;    APPLICANT:  Goldberg, Edward B.
;    TITLE OF INVENTION:  MATERIALS FOR THE PRODUCTION OF
;    TITLE OF INVENTION:  NANOMETER STRUCTURES AND USE THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie and Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  US
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/542,003
;      FILING DATE:  13-OCT-1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  8471-0005-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66441 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8855 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacteriophage T4
```

```
;    IMMEDIATE SOURCE:
;      CLONE:  TAIL FIBER GENES
US-08-542-003-1

Alignment Scores:
Pred. No.:                271             Length:         8855
Score:                    71.50           Matches:        25
Percent Similarity:       46.05%          Conservative:   10
Best Local Similarity:    32.89%          Mismatches:     24
Query Match:              8.43%           Indels:         17
DB:                       2               Gaps:           5

US-09-900-575-29_COPY_26_186 (1-161) x US-08-542-003-1 (1-8855)

Qy      42 ValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThr 61
           ||||||          ||||||   ::::::         :::   ||||||   ||||||
Db    6757 GTACTACGGAAACTGTCAGGAGTAATAGAA------GCAACAGAATATCCACCGCCAACT 6704

Qy      62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysPro---------Trp 78
              |||   ||||||      :::|||           |||||||||
Db    6703 AGATCAAATACACCAGTTTTTTTGTAT-----------GACAAGCCAGTTACAGTGTCG 6656

Qy      79 ProValAlaLeuTyrLeuThrProVal-----------------SerSerAlaGlyGly 92
           |||         ||||||:::|||:::                 |||      |||
Db    6655 CCGAGAACAAGATACTTATCGCCCATTACACCGACGTTGCCATAATCACTAGACGGGAAA 6596

Qy      93 LeuValIleLysAla------GlySerLeuIleAlaValLeuIleLeu 106
              |||:::           |||:::   ::::::||||||
Db    6595 TTTGTAGTGAATGTACCTAATGCAAGGCTATTACGGATAATCATTCTG 6548

RESULT 29
US-08-322-760A-1/c
; Sequence 1, Application US/08322760A
; Patent No. 5877279
;  GENERAL INFORMATION:
;    APPLICANT:  Goldberg, Edward B.
;    TITLE OF INVENTION:  MATERIALS FOR THE PRODUCTION OF
;    TITLE OF INVENTION:  NANOMETER STRUCTURES AND USE THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie and Edmonds
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  US
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/322,760A
;      FILING DATE:  13-OCT-1994
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  8471-0003-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66441 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8855 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Bacteriophage T4
;    IMMEDIATE SOURCE:
;      CLONE:  TAIL FIBER GENES
US-08-322-760A-1

Alignment Scores:
Pred. No.:                271             Length:         8855
Score:                    71.50           Matches:        25
Percent Similarity:       46.05%          Conservative:   10
Best Local Similarity:    32.89%          Mismatches:     24
Query Match:              8.43%           Indels:         17
DB:                       2               Gaps:           5

US-09-900-575-29_COPY_26_186 (1-161) x US-08-322-760A-1 (1-8855)

Qy      42 ValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThr 61
           ||||||          ||||||   ::::::         :::   ||||||   ||||||
Db    6757 GTACTACGGAAACTGTCAGGAGTAATAGAA------GCAACAGAATATCCACCGCCAACT 6704

Qy      62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysPro---------Trp 78
              |||   ||||||      :::|||           |||||||||
Db    6703 AGATCAAATACACCAGTTTTTTTGTAT-----------GACAAGCCAGTTACAGTGTCG 6656

Qy      79 ProValAlaLeuTyrLeuThrProVal-----------------SerSerAlaGlyGly 92
           |||         ||||||:::|||:::                 |||      |||
Db    6655 CCGAGAACAAGATACTTATCGCCCATTACACCGACGTTGCCATAATCACTAGACGGGAAA 6596

Qy      93 LeuValIleLysAla------GlySerLeuIleAlaValLeuIleLeu 106
              |||:::           |||:::   ::::::||||||
Db    6595 TTTGTAGTGAATGTACCTAATGCAAGGCTATTACGGATAATCATTCTG 6548

RESULT 30
US-09-236-949-1/c
; Sequence 1, Application US/09236949
; Patent No. 6437112
;  GENERAL INFORMATION:
;       APPLICANT: Goldberg, Edward B.
;       TITLE OF INVENTION: MATERIALS FOR THE PRODUCTION OF
;                           NANOMETER STRUCTURES AND USE THEREOF
;       NUMBER OF SEQUENCES: 6
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pennie and Edmonds
;            STREET: 1155 Avenue of the Americas
;            CITY: New York
;            STATE: New York
;            COUNTRY: US
;            ZIP: 10036
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/236,949
;            FILING DATE: 25-Jan-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/542,003
;            FILING DATE: 13-OCT-1995
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Misrock, S. Leslie
;            REGISTRATION NUMBER: 18,872
;            REFERENCE/DOCKET NUMBER: 8471-0005-999
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (212) 790-9090
;            TELEFAX: 212-869-8864
;            TELEX: 66441 PENNIE
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 8855 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
```

```
;       MOLECULE TYPE: DNA (genomic)
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       ORIGINAL SOURCE:
;            ORGANISM: Bacteriophage T4
;       IMMEDIATE SOURCE:
;            CLONE: TAIL FIBER GENES
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-236-949-1

Alignment Scores:
Pred. No.:                271           Length:          8855
Score:                    71.50         Matches:         25
Percent Similarity:       46.05%        Conservative:    10
Best Local Similarity:    32.89%        Mismatches:      24
Query Match:              8.43%         Indels:          17
DB:                       4             Gaps:            5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-236-949-1 (1-8855)

Qy      42 ValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThr 61
           ||||||         ||||||  ::::::        :::   ||||||   ||||||
Db    6757 GTACTACGGAAACTGTCAGGAGTAATAGAA------GCAACAGAATATCCACCGCCAACT 6704

Qy      62 ThrSerGluThrProArgValValTyrAsnSerArgThrAspLysPro---------Trp 78
             |||   ||||||      :::|||            |||||||||
Db    6703 AGATCAAATACACCAGTTTTTTGTAT-----------GACAAGCCAGTTACAGTGTCG 6656

Qy      79 ProValAlaLeuTyrLeuThrProVal-----------------SerSerAlaGlyGly 92
           |||         ||||||:::|||:::                 |||      |||
Db    6655 CCGAGAACAAGATACTTATCGCCCATTACACCGACGTTGCCATAATCACTAGACGGGAAA 6596

Qy      93 LeuValIleLysAla------GlySerLeuIleAlaValLeuIleLeu 106
             |||:::             |||:::   ::::::||||||
Db    6595 TTTGTAGTGAATGTACCTAATGCAAGGCTATTACGGATAATCATTCTG 6548

RESULT 31
US-08-743-637B-27/c
; Sequence 27, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  PICARD, Francois J.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;    TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;    TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
;    TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 EAST WISCONSIN AVENUE
;      CITY:  MILWAUKEE
;      STATE:  WISCONSIN
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/743,637B
;      FILING DATE:  04-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,840
;      FILING DATE:  11-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414)277-5591
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Haemophilus influenzae
US-08-743-637B-27

Alignment Scores:
Pred. No.:                282           Length:          9100
Score:                    71.50         Matches:         39
Percent Similarity:       44.53%        Conservative:    22
Best Local Similarity:    28.47%        Mismatches:      57
Query Match:              8.43%         Indels:          19
DB:                       2             Gaps:            8

US-09-900-575-29_COPY_26_186 (1-161) x US-08-743-637B-27 (1-9100)

Qy      34 GlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSer 53
           ::::::|||      |||         |||         ||||||      :::|||
Db    884 GAAAAAGGCTTTTTTTAT---------TTGGATACGCACTCTGGTGTGGGGCGTTATCGT 834

Qy      54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsn------ 71
             ||||||                  |||   |||           :::
Db    833 TTATCCTCAAATGAATCAGAAAAAACGGGGGAATATAAAGAAGGTATTGGACGCCTGTGG 774

Qy      72 SerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer------Ser 89
             :::||||||  |||      :::|||   |||:::      :::         :::
Db    773 GATCAAACAGATTTACCCGAAGATATTGCTCGTTATGTAAAAATGATCAAAAAACTCAAT 714

Qy      90 AlaGlyGlyLeuValIleLys------AlaGlySer---LeuIleAlaValLeuIleLeu 106
             ||||||      ::::::      |||||||||   |||||||||      :::|||
Db    713 TATGGTGGCAAAGAACTACGTTATTACGCGGGTTCTCCATTAATTGCCGCGGAATTGTTG 654

Qy    107 ArgGlnThrAsn---------------AsnTyrAsnSerAspAspPheGlnPheValTrp 121
           |||       :::                :::         |||:::      :::
Db    653 CGCTCACAAGATCGCGCACTATTGACCGAGCTTCATCCTAGCGATTATCCAATTCTTCGC 594

Qy    122 AsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAlaArg 141
           |||   ::::::::::::      |||      |||      ||||||          :::
Db    593 AATAATTTTAGCGACGACAAAAATGTC------ACCGTAAAATGTGACAATGGCTTTCAA 540

Qy    142 AspValThrValThrLeuPro------AspTyrArgGlySerValProIle 156
           |||      |||||||||      :::   ||||||   |||   |||
Db    539 CAAGTCAAAGCAACGCTTCCGCCAAAAGAACGCCGAGGCTTAGTACTCATC 489

RESULT 32
US-08-526-840B-27/c
; Sequence 27, Application US/08526840B
; Patent No. 6001564
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIFIC AND UNIVERSAL PROBES AND
;    TITLE OF INVENTION:  AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
;    TITLE OF INVENTION:  COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
;    TITLE OF INVENTION:  FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
;    NUMBER OF SEQUENCES:  177
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  Wisconsin
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/526,840B
;       FILING DATE:  11-SEP-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/304,732
;       FILING DATE:  12-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BAKER, Jean C.
;       REGISTRATION NUMBER:  35,433
;       REFERENCE/DOCKET NUMBER:  850586.90012
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (414) 277-5000
;       TELEFAX:  (414) 277-5591
;   INFORMATION FOR SEQ ID NO:  27:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9100 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Haemophilus influenzae
US-08-526-840B-27

Alignment Scores:
Pred. No.:                282             Length:         9100
Score:                    71.50           Matches:        39
Percent Similarity:       44.53%          Conservative:   22
Best Local Similarity:    28.47%          Mismatches:     57
Query Match:              8.43%           Indels:         19
DB:                       3               Gaps:           8

US-09-900-575-29_COPY_26_186 (1-161) x US-08-526-840B-27 (1-9100)

Qy      34 GlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSer 53
           :::::::|||      |||         |||         ||||||      :::|||
Db     884 GAAAAAGGCTTTTTTTAT---------TTGGATACGCACTCTGGTGTGGGGCGTTATCGT 834

Qy      54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsn------ 71
             ||||||                |||   |||                 :::
Db     833 TTATCCTCAAATGAATCAGAAAAAACGGGGGAATATAAAGAAGGTATTGGACGCCTGTGG 774

Qy      72 SerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer------Ser 89
           :::|||||||  |||     :::|||   |||:::      :::         :::
Db     773 GATCAAACAGATTTACCCGAAGATATTGCTCGTTATGTAAAAATGATCAAAAAACTCAAT 714

Qy      90 AlaGlyGlyLeuValIleLys------AlaGlySer---LeuIleAlaValLeuIleLeu 106
           ||||||      ::::::      |||||||||   |||||||||      :::|||
Db     713 TATGGTGGCAAAGAACTACGTTATTACGCGGGTTCTCCATTAATTGCCGCGGAATTGTTG 654

Qy     107 ArgGlnThrAsn--------------AsnTyrAsnSerAspAspPheGlnPheValTrp 121
           |||      :::                :::      |||:::      :::
Db     653 CGCTCACAAGATCGCGCACTATTGACCGAGCTTCATCCTAGCGATTATCCAATTCTTCGC 594

Qy     122 AsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAlaArg 141
           |||   ::::::::::::      |||      |||      ||||||         :::
Db     593 AATAATTTTAGCGACGACAAAAATGTC------ACCGTAAAATGTGACAATGGCTTTCAA 540

Qy     142 AspValThrValThrLeuPro------AspTyrArgGlySerValProIle 156
           |||      |||||||||      :::   ||||||   |||   |||
Db     539 CAAGTCAAAGCAACGCTTCCGCCAAAAGAACGCCGAGGCTTAGTACTCATC 489

RESULT 33
US-08-370-319C-2/c
; Sequence 2, Application US/08370319C
; Patent No. 5856091
;  GENERAL INFORMATION:
;    APPLICANT:  Brichard, Vincent; Van Pel, Aline;
;    APPLICANT:  Traversari, Catia; W lfel, Thomas; Coulie, Pierre;
;    APPLICANT:  Boon-Falleur, Thierry; De Plaen, Etienne
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
;    TITLE OF INVENTION:  TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT LEAST ONE
;    TITLE OF INVENTION:  REJECTION ANTIGEN PRESENTED BY HLA-A2
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/370,319C
;      FILING DATE:  10-JANUARY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/272,351
;      FILING DATE:  8-JULY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/032,978
;      FILING DATE:  18-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 5856091man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5377.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9421 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  Following position there is an
;      OTHER INFORMATION:  unsequenced portion of from 4.7 to 5.3
;      OTHER INFORMATION:  kilobases
US-08-370-319C-2

Alignment Scores:
Pred. No.:                297             Length:         9421
Score:                    71.50           Matches:        20
Percent Similarity:       43.94%          Conservative:   9
Best Local Similarity:    30.30%          Mismatches:     20
Query Match:              8.43%           Indels:         17
DB:                       2               Gaps:           3

US-09-900-575-29_COPY_26_186 (1-161) x US-08-370-319C-2 (1-9421)

Qy      29 AspTyrValThrLeuGlnArg----------------------GlySerAlaTyrGly 40
           :::|||:::|||||||||||||                    |||         |||
Db    6935 AATTATCTGACTCTCCAGAGATTCAGAGGAAAGAACTGGAAAGAGGGCCCTGGACTTGGA 6876

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhe--- 59
           |||:::      :::      ||||||:::             :::   |||||||||
Db    6875 GGGATT------TCAGCAGTGGGAACTTTACCAACCATCAAGGCTCTGTATCCATTTCGT 6822

Qy      60 -----------------ProThrThrSerGluThrProArgValValTyrAsnSerArg 73
                           |||   :::|||:::||||||||||:::
Db    6821 CTTCTACAATACCAACAGCCGATGAGCAGTAAGACTCCCAGGATCACTGTCAGGATGCCG 6762

Qy      74 ThrAspLysProTrpPro 79
                 ||||||
Db    6761 ATCCCAGCGGCCTGGCCC 6744
```

```
RESULT 34
US-09-224-834-2/c
; Sequence 2, Application US/09224834
; Patent No. 6201111
;  GENERAL INFORMATION:
;    APPLICANT:  Brichard, Vincent; Van Pel, Aline;
;    APPLICANT:  Traversari, Catia; W lfel, Thomas; Coulie, Pierre;
;    APPLICANT:  Boon-Falleur, Thierry; De Plaen, Etienne
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
;    TITLE OF INVENTION:  TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT LEAST ONE TU
;    TITLE OF INVENTION:  REJECTION ANTIGEN PRESENTED BY HLA-A2
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Felfe & Lynch
;      STREET:  805 Third Avenue
;      CITY:  New York City
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 5.25 inch, 360 kb storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  Wordperfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/224,834
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/370,319
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/032,978
;      FILING DATE:  18-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hanson, No. 6201111man D.
;      REGISTRATION NUMBER:  30,946
;      REFERENCE/DOCKET NUMBER:  LUD 5377.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 688-9200
;      TELEFAX:  (212) 838-3884
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9421 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  Following position there is an
;      OTHER INFORMATION:  unsequenced portion of from 4.7 to 5.3
;      OTHER INFORMATION:  kilobases
US-09-224-834-2

Alignment Scores:
Pred. No.:                297              Length:          9421
Score:                    71.50            Matches:         20
Percent Similarity:       43.94%           Conservative:    9
Best Local Similarity:    30.30%           Mismatches:      20
Query Match:              8.43%            Indels:          17
DB:                       4                Gaps:            3

US-09-900-575-29_COPY_26_186 (1-161) x US-09-224-834-2 (1-9421)

Qy      29 AspTyrValThrLeuGlnArg----------------------GlySerAlaTyrGly 40
           :::|||:::|||||||||||                      |||         |||
Db    6935 AATTATCTGACTCTCCAGAGATTCAGAGGAAAGAACTGGAAAGAGGGCCCTGGACTTGGA 6876

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhe--- 59
           |||:::      :::      ||||||:::           :::   |||||||||
Db    6875 GGGATT------TCAGCAGTGGGAACTTTACCAACCATCAAGGCTCTGTATCCATTTCGT 6822

Qy      60 -----------------ProThrThrSerGluThrProArgValValTyrAsnSerArg 73
                           |||   :::|||:::||||||||||:::

Db    6821 CTTCTACAATACCAACAGCCGATGAGCAGTAAGACTCCCAGGATCACTGTCAGGATGCCG 6762

Qy      74 ThrAspLysProTrpPro 79
                   ||||||
Db    6761 ATCCCAGCGGCCTGGCCC 6744

RESULT 35
US-09-110-959A-1
; Sequence 1, Application US/09110959A
; Patent No. 6268197
; GENERAL INFORMATION:
;  APPLICANT:  Schulein, Martin
;  APPLICANT:  Outtrup, Helle
;  APPLICANT:  Jorgensen, Per Lina
;  APPLICANT:  Bjornvad, Mads Eskelund
;  TITLE OF INVENTION:  Alkaline Xyloglucanase
;  FILE REFERENCE:  5206.200-US
;  CURRENT APPLICATION NUMBER: US/09/110,959A
;  CURRENT FILING DATE:  1998-07-07
;  PRIOR APPLICATION NUMBER:  0822/97
;  PRIOR FILING DATE:  1997-07-07
;  PRIOR APPLICATION NUMBER:  1213/97
;  PRIOR FILING DATE:  1997-10-24
;  PRIOR APPLICATION NUMBER:  60/054,039
;  PRIOR FILING DATE:  1997-07-28
;  PRIOR APPLICATION NUMBER:  60/063,694
;  PRIOR FILING DATE:  1997-10-28
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 786
;   TYPE: DNA
;   ORGANISM: Bacillus licheniformis ATCC 14580
US-09-110-959A-1

Alignment Scores:
Pred. No.:                8.81             Length:          786
Score:                    71.00            Matches:         34
Percent Similarity:       36.54%           Conservative:    23
Best Local Similarity:    21.79%           Mismatches:      65
Query Match:              8.37%            Indels:          34
DB:                       4                Gaps:            5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-110-959A-1 (1-786)

Qy       4 AsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHisAsn------ 21
           ::::::     |||          :::   :::      |||:::   ||||||
Db     373 TCGATCAGCGCAAACGGCACATACAATGCCGCATATGACATTTGGCTCCACAATACAAAC 432

Qy      22 ------AspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyr 39
                 :::          ||||||   :::   :::          :::
Db     433 AAGGCGAGCTGGGATTCGGCTCCAACCGATGAGATTATGATCTGGCTCAATAACACAAAC 492

Qy      40 GlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhe 59
             |||         ::::::      ||||||         |||   |||:::
Db     493 GCCGGACCTGCCGGTTCCTATGTCGAAACTGTATCGATTGGCGGGCACAGTTGGAAA--- 549

Qy      60 ProThrThrSerGluThrProArgValValTyrAsnSerArgThrAsp----------- 75
                                    ||||||           |||
Db     550 -------------------------GTATATAAAGGCTATATTGATGCTGGAGGCGGC 582

Qy      76 LysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIle 95
           |||   |||   |||      ::::::
Db     583 AAAGGGTGGAACGTGTTTTCATTTATCAGA---------------------------- 612

Qy      96 LysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAsp 115
             |||   :::      |||   |||   :::|||                |||||   :::|||
Db     613 ACAGCAAACACCCAAAGTGCGAACCTGAATATTCGGGATTTCACGAATTATCTTGCCGAC 672

Qy     116 AspPheGln-----------------PheValTrpAsnIleTyrAlaAsnAsnAspVal 129
             |||                  :::|||   ::::::           :::|||
Db     673 TCCAAACAGTGGCTTTCCAAAACAAAGTATGTCAGCAGTGTGGAATTCGGTACTGAAGTT 732
```

```
Qy     130 ValValProThrGlyGlyCysAspValSerAlaArgAspValThrVal 145
                 ||||||      :::::::|||      ||||||||||||
Db     733 TTCGGAGGCACAGGACAAATTAATATTTCCAATTGGGACGTAACGGTC 780

RESULT 36
US-09-071-035-263/c
; Sequence 263, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  263:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1463 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-071-035-263

Alignment Scores:
Pred. No.:                22              Length:          1463
Score:                    71.00           Matches:         35
Percent Similarity:       37.32%          Conservative:    18
Best Local Similarity:    24.65%          Mismatches:      40
Query Match:              8.37%           Indels:          49
DB:                       4               Gaps:            6

US-09-900-575-29_COPY_26_186 (1-161) x US-09-071-035-263 (1-1463)

Qy      19 CysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAla 38
           |||   |||      |||      ||||||
Db     757 TGTTGCAATCACGTTCCCATCTGCATCACG--------------------------- 728

Qy      39 TyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrPro 58
                        ||||||   |||:::::::::   ||||||:::   :::
Db     727 -----------------AATTTCGATGGTAGCATTCGGATCAGCGGTGCCTGTCACCTG 686

Qy      59 Phe-ProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTr 78
              |||   |||
Db     685 ATAACCAGCAACTGAA---------------------------------------- 670

Qy      78 pProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGl 98
                 |||   :::||||||||||:::               |||      |||||
Db     669 ----------TTACCTGTTACTCCTGTTACAGTTGGTGCTGTAACCGTGGCATCAGCAGG 620

Qy      98 ySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGl 118
           |:::      :::|||      |||            ::::::   :::      |||
Db     619 AACAAGGACTTCAGTCGGTTGACTT-----------TCATTTCCCGCGTTATCCTTTCC 572

Qy     118 nPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspVa 138
                  |||:::      :::      :::||||||||||   |||
Db     571 AATCACTGTTAACGTA------TCCCCAGGATTGGTCGTTCCCGTTGGT---------- 529

Qy     138 lSerAlaArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLe 158
                       |||||||||      |||   :::         ||||||:::|||::
Db     528 ----------AGCGTTACAGTAAAATCACCA---TTCGCATCAGAGGTACCTGTACCAAT 482

Qy     158 uThr 159
           :|||
Db     481 GACT 478

RESULT 37
US-09-071-035-257/c
; Sequence 257, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  257:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4977 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-071-035-257

Alignment Scores:
Pred. No.:                134             Length:          4977
Score:                    71.00           Matches:         35
Percent Similarity:       37.32%          Conservative:    18
Best Local Similarity:    24.65%          Mismatches:      40
Query Match:              8.37%           Indels:          49
DB:                       4               Gaps:            6

US-09-900-575-29_COPY_26_186 (1-161) x US-09-071-035-257 (1-4977)

Qy      19 CysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAla 38
           |||   |||      |||      ||||||
Db    2580 TGTTGCAATCACGTTCCCATCTGCATCACG--------------------------- 2551

Qy      39 TyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrPro 58
                        ||||||   |||:::::::::   ||||||:::   :::
```

```
Db   2550 ------------------AATTTCGATGGTAGCATTCGGATCAGCGGTGCCTGTCACCTG 2509

Qy     59 Phe-ProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTr 78
             |||   |||
Db   2508 ATAACCAGCAACTGAA---------------------------------------------- 2493

Qy     78 pProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGl 98
               |||   :::|||||||||:::                |||      |||||
Db   2492 ----------TTACCTGTTACTCCTGTTACAGTTGGTGCTGTAACCGTGGCATCAGCAGG 2443

Qy     98 ySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGl 118
          |:::      :::|||      |||           ::::::   :::      |||
Db   2442 AACAAGGACTTCAGTCGGTTGACTT-----------TCATTTCCCGCGTTATCCTTTCC 2395

Qy    118 nPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspVa 138
                 |||:::      :::       :::|||||||||   |||
Db   2394 AATCACTGTTAACGTA------TCCCCAGGATTGGTCGTTCCCGTTGGT---------- 2352

Qy    138 lSerAlaArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLe 158
                  |||||||||      |||   :::         |||||:::|||::
Db   2351 ----------AGCGTTACAGTAAAATCACCA---TTCGCATCAGAGGTACCTGTACCAAT 2305

Qy    158 uThr 159
          :|||
Db   2304 GACT 2301
```

RESULT 38
US-09-071-035-261/c
; Sequence 261, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 261:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-071-035-261

Alignment Scores:

Pred. No.:	134	Length:	4977
Score:	71.00	Matches:	35
Percent Similarity:	37.32%	Conservative:	18
Best Local Similarity:	24.65%	Mismatches:	40
Query Match:	8.37%	Indels:	49
DB:	4	Gaps:	6

US-09-900-575-29_COPY_26_186 (1-161) x US-09-071-035-261 (1-4977)

```
Qy     19 CysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAla 38
          |||   |||       |||       ||||||
Db   2580 TGTTGCAATCACGTTCCCATCTGCATCACG---------------------------- 2551

Qy     39 TyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrPro 58
                              ||||||   |||:::::::::   |||||||:::   :::
Db   2550 -----------------AATTTCGATGGTAGCATTCGGATCAGCGGTGCCTGTCACCTG 2509

Qy     59 Phe-ProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTr 78
             |||   |||
Db   2508 ATAACCAGCAACTGAA---------------------------------------------- 2493

Qy     78 pProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGl 98
               |||   :::||||||||||:::                |||      |||||
Db   2492 ----------TTACCTGTTACTCCTGTTACAGTTGGTGCTGTAACCGTGGCATCAGCAGG 2443

Qy     98 ySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGl 118
          |:::      :::|||      |||           ::::::   :::      |||
Db   2442 AACAAGGACTTCAGTCGGTTGACTT-----------TCATTTCCCGCGTTATCCTTTCC 2395

Qy    118 nPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspVa 138
                 |||:::      :::       :::|||||||||   |||
Db   2394 AATCACTGTTAACGTA------TCCCCAGGATTGGTCGTTCCCGTTGGT---------- 2352

Qy    138 lSerAlaArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLe 158
                  |||||||||      |||   :::         |||||:::|||::
Db   2351 ----------AGCGTTACAGTAAAATCACCA---TTCGCATCAGAGGTACCTGTACCAAT 2305

Qy    158 uThr 159
          :|||
Db   2304 GACT 2301
```

RESULT 39
US-09-071-035-265/c
; Sequence 265, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 265:
; SEQUENCE CHARACTERISTICS:

```
;       LENGTH:  4977 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-09-071-035-265

Alignment Scores:
Pred. No.:                134             Length:         4977
Score:                    71.00           Matches:        35
Percent Similarity:       37.32%          Conservative:   18
Best Local Similarity:    24.65%          Mismatches:     40
Query Match:              8.37%           Indels:         49
DB:                       4               Gaps:           6

US-09-900-575-29_COPY_26_186 (1-161) x US-09-071-035-265 (1-4977)

Qy      19 CysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAla 38
           |||   |||       |||       ||||||
Db    2580 TGTTGCAATCACGTTCCCATCTGCATCACG---------------------------- 2551

Qy      39 TyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrPro 58
                       ||||||   |||:::::::::  ||||||:::   :::
Db    2550 -----------------AATTTCGATGGTAGCATTCGGATCAGCGGTGCCTGTCACCTG 2509

Qy      59 Phe-ProThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTr 78
              |||   |||
Db    2508 ATAACCAGCAACTGAA---------------------------------------------- 2493

Qy      78 pProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGl 98
                |||   :::||||||||||:::                 |||       |||||
Db    2492 ----------TTACCTGTTACTCCTGTTACAGTTGGTGCTGTAACCGTGGCATCAGCAGG 2443

Qy      98 ySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGl 118
           |:::      :::|||      |||             ::::::  ':::      |||
Db    2442 AACAAGGACTTCAGTCGGTTGACTT-----------TCATTTCCCGCGTTATCCTTTCC 2395

Qy     118 nPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspVa 138
                |||:::      :::      :::||||||||||   |||
Db    2394 AATCACTGTTAACGTA------TCCCCAGGATTGGTCGTTCCCGTTGGT---------- 2352

Qy     138 lSerAlaArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLe 158
                     |||||||||      |||  :::          ||||||:::|||::
Db    2351 ----------AGCGTTACAGTAAAATCACCA---TTCGCATCAGAGGTACCTGTACCAAT 2305

Qy     158 uThr 159
           :|||
Db    2304 GACT 2301

RESULT 40
US-08-961-527-71
; Sequence 71, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  71:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  32768 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-71

Alignment Scores:
Pred. No.:                2.49e+03        Length:         32768
Score:                    70.50           Matches:        29
Percent Similarity:       34.40%          Conservative:   14
Best Local Similarity:    23.20%          Mismatches:     43
Query Match:              8.31%           Indels:         39
DB:                       4               Gaps:           4

US-09-900-575-29_COPY_26_186 (1-161) x US-08-961-527-71 (1-32768)

Qy       7 GlnAsnLeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThr 26
           |||   |||:::   |||||||||       :::|||
Db   20684 CAACAACTAATCAATGATTTAAGTAAATCTCTTTTT---------------------- 20719

Qy      27 IleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeu--------- 43
                                        |||       ||||||      |||
Db   20720 --------------------------GGCGGTGGGTATGGTGATAAACTCTTTCCTCCG 20752

Qy      44 --------------SerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrPro 58
                          ||||||:::|||   :::   |||   |||       |||
Db   20753 ACGATACAAGTCAATCCAAATTTTACAGGAGCTATTTCTTATCAAGGATTGGATTATGTC 20812

Qy      59 Phe-----------------ProThrThrSerGluThrProArgValValTyrAsnSer 72
                                :::::                  ||||||
Db   20813 AGTTTGGAAGGTGAGTTTGGGCAAGATTTTGCCCAGCTTGCCTATTGGGCTTATAATATT 20872

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
                  |||       |||:::   |||:::|||
Db   20873 ATGGTGCAAAAAACACTCCCTATTGAGTTGTGGCTTGAA---------------------- 20911

Qy      93 LeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyr 112
                  |||   |||:::            |||::::::|||:::      :::
Db   20912 ---TATGAGAAGGAAGGCAATTGTGACTTTCGTTTAGTAATCCGTAAAATGTGGAGTGGG 20968

Qy     113 AsnSerAspAspPhe 117
           :::   |||||||||
Db   20969 TCTGTTGATGATTTC 20983

RESULT 41
US-09-295-186-9
; Sequence 9, Application US/09295186B
; Patent No. 6127137
; GENERAL INFORMATION:
;  APPLICANT:  Hasida, Miyoko
;  APPLICANT:  Tsutsumi, No. 6127137iko
;  APPLICANT:  Halkier, Torben
;  APPLICANT:  Stringer, Mary Ann
;  TITLE OF INVENTION:  An Acidic Phospholipase, Production, and
;  TITLE OF INVENTION:  Methods of Using Thereof (As Amended)
;  FILE REFERENCE:  4953.204-US
;  CURRENT APPLICATION NUMBER: US/09/295,186B
;  CURRENT FILING DATE:  1999-04-20
;  PRIOR APPLICATION NUMBER:  1215/96
;  PRIOR FILING DATE:  1996-10-31
;  PRIOR APPLICATION NUMBER:  PCT/DK97/00490
;  PRIOR FILING DATE:  1997-10-30
```

```
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 9
;   LENGTH: 1870
;   TYPE: DNA (genomic)
;   ORGANISM: Hyphozyma sp. CBS 648.91
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (214)...(1869)
;   NAME/KEY: mat_peptide
;   LOCATION: (442)...(1869)
US-09-295-186-9

Alignment Scores:
Pred. No.:                42.2          Length:          1870
Score:                    70.00         Matches:         33
Percent Similarity:       35.51%        Conservative:    16
Best Local Similarity:    23.91%        Mismatches:      45
Query Match:              8.25%         Indels:          44
DB:                       3             Gaps:            7

US-09-900-575-29_COPY_26_186 (1-161) x US-09-295-186-9 (1-1870)

Qy      23 TyrProGluThrIleThrAspTyrValThrLeuGln----------------ArgGly 36
           |||      ::::::|||||||||      |||                     |||
Db    1006 TACCAGACCAGTCTCACCGACTACTTTGGCCTCTCGATCGCCGACAAGATTCTCAACGGC 1065

Qy      37 SerAlaTyrGlyGlyValLeu-----------SerAsnPheSerGlyThrValLysTyr 52
           |||   ||||||                   |||:::      |||   |||:::
Db    1066 TCCATGTACGGCAACAAGTTCAGCGTCGAGTGGAGCGACGTCAAGAATACGTCCAAGTTC 1125

Qy      53 SerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSer 72
           :::   :::|||   |||||||||                 ::::::   :::
Db    1126 ACCGATGCCTCCATGCCGTTCCCCATC----------------ATTATTGCCGACGAG 1167

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGly 92
           |||
Db    1168 CGC-------------------------------------------------------- 1170

Qy      93 LeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyr 112
              :::   |||   ||||||         |||   |||   |||      :::
Db    1171 ---------GAGCCCGGCGAGCTCATC---------ATCCCGCGCAACACCACCATCTGG 1212

Qy     113 AsnSerAspAspPheGlnPhe---ValTrpAsnIleTyrAlaAsnAsnAspValValVal 131
              :::   ::::::|||      ||||||         |||            :::
Db    1213 GAGTTCAACCCGTACGAGTTCGGTTCTTGGAAC------CCCAATGTTTCGGCTTTCATC 1266

Qy     132 ProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAsp 149
           |||                 |||      :::   |||      |||||||||
Db    1267 CCCATCGAGATCCTCGGCTCGAGTCTGGACAACGGCACCAGCGTCCTGCCCGAC 1320

RESULT 42
US-08-976-259-85/c
; Sequence 85, Application US/08976259
; Patent No. 6316609
;  GENERAL INFORMATION:
;    APPLICANT:  Dillon, Patrick J.
;    APPLICANT:  Choi, Gil H.
;    APPLICANT:  Welch, Rodney A.
;    TITLE OF INVENTION:  Nucleotide Sequence of Escherichia coli
; Patent No. 6316609
;    NUMBER OF SEQUENCES:  142
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Ave, N.W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/976,259
;      FILING DATE:  Herewith
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/031,626 AND US 60/061,953
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Steffe, Eric K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0740002/EKS/CBM
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  85:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-976-259-85

Alignment Scores:
Pred. No.:                521           Length:          9319
Score:                    69.50         Matches:         32
Percent Similarity:       40.00%        Conservative:    28
Best Local Similarity:    21.33%        Mismatches:      71
Query Match:              8.20%         Indels:          19
DB:                       4             Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-08-976-259-85 (1-9319)

Qy      26 ThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsn 45
           |||:::::::::      |||:::|||   |||         |||||||||      |||
Db    7073 ACCCTGAGTAACCAGGGCACGATACAGGGTGGT---------GGCGGGGTTTCCCTGAAC 7023

Qy      46 PheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThr 65
              :::     ::::::      |||                              :::   |||
Db    7022 GCCACTGACCGTCTGCAGAACGACGGCAAAATCCTCTCCGGCAGTAACCTCACGCTGACG 6963

Qy      66 ProArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThr 85
              :::|||:::   |||:::   :::                          |||   |||
Db    6962 GCGCAGGTGCTGGCGAACACCGGCAGCGGACTGGTACAGGCTGCCACCCTGCTGCTGGAT 6903

Qy      86 ProValSerSerAla---GlyGlyLeuValIleLysAlaGlySerLeu---IleAlaVal 103
              |||::::::      ||||||   |||:::      ||||||      :::
Db    6902 GTGGTGAATACTGTCAACGGCGGACGCGTACTTGCCACCGGCAGTGCCGACGTTAAAGGA 6843

Qy     104 LeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIle 123
              |||      |||                  |||          ::::::
Db    6842 ACCACGCTGAATAATACCGGTACGCTTCAGGGTGCGGACCTGCTGGTGAATTACCACACA 6783

Qy     124 TyrAlaAsnAsnAspValValValProThrGlyGlyCysAspVal--------------- 138
           ::::::|||:::      ::::::   |||   |||      |||
Db    6782 TTCAGCAACAGCGGTACCCTGCTGGGAACCTCCGGGCTTGGCGTCAAGGGCAGTTCACTG 6723

Qy     139 --------------------------SerAlaArgAspValThrValThrLeuProAsp 149
                                       ||||||   ::::::   :::         |||
Db    6722 CTGCAAAATGGTACAGGGCGGCTGTACAGTGCAGGCAACCTGCTGCTTGACGCTCAGGAC 6663

Qy     150 TyrArgGlySerValProIleProLeuThr 159
           :::   |||         :::      |||
Db    6662 TTCAGTGGTCAGGGGCAGGTGGTGGCCACC 6633

RESULT 43
US-08-976-259-14
; Sequence 14, Application US/08976259
; Patent No. 6316609
;  GENERAL INFORMATION:
;    APPLICANT:  Dillon, Patrick J.
;    APPLICANT:  Choi, Gil H.
;    APPLICANT:  Welch, Rodney A.
;    TITLE OF INVENTION:  Nucleotide Sequence of Escherichia coli
```

```
; Patent No. 6316609
;    NUMBER OF SEQUENCES:  142
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Ave, N.W., Suite 600
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/976,259
;      FILING DATE:  Herewith
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/031,626 AND US 60/061,953
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Steffe, Eric K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0740002/EKS/CBM
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  22671 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-976-259-14

Alignment Scores:
Pred. No.:                1.93e+03        Length:          22671
Score:                    69.50           Matches:         32
Percent Similarity:       40.00%          Conservative:    28
Best Local Similarity:    21.33%          Mismatches:      71
Query Match:              8.20%           Indels:          19
DB:                       4               Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-08-976-259-14 (1-22671)

Qy      26 ThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsn 45
           |||::::::::::      |||:::|||   |||         |||||||||      |||
Db   16130 ACCCTGAGTAACCAGGGCACGATACAGGGTGGT---------GGCGGGGTTTCCCTGAAC 16180

Qy      46 PheSerGlyThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThr 65
             :::     ::::::      |||                              :::   |||
Db   16181 GCCACTGACCGTCTGCAGAACGACGGCAAAATCCTCTCCGGCAGTAACCTCACGCTGACG 16240

Qy      66 ProArgValValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThr 85
             :::|||:::   |||:::   :::                        |||   |||
Db   16241 GCGCAGGTGCTGGCGAACACCGGCAGCGGACTGGTACAGGCTGCCACCCTGCTGCTGGAT 16300

Qy      86 ProValSerSerAla---GlyGlyLeuValIleLysAlaGlySerLeu---IleAlaVal 103
             |||::::::      ||||||   |||:::      ||||||      :::
Db   16301 GTGGTGAATACTGTCAACGGCGGACGCGTACTTGCCACCGGCAGTGCCGACGTTAAAGGA 16360

Qy     104 LeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheValTrpAsnIle 123
           |||      |||                  |||                 ::::::
Db   16361 ACCACGCTGAATAATACCGGTACGCTTCAGGGTGCGGACCTGCTGGTGAATTACCACACA 16420

Qy     124 TyrAlaAsnAsnAspValValValProThrGlyGlyCysAspVal-------------- 138
           ::::::|||:::      ::::::   |||  |||      |||
Db   16421 TTCAGCAACAGCGGTACCCTGCTGGGAACCTCCGGGCTTGGCGTCAAGGGCAGTTCACTG 16480

Qy     139 --------------------------SerAlaArgAspValThrValThrLeuProAsp 149
                                       ||||||   ::::::   :::         |||
Db   16481 CTGCAAAATGGTACAGGGCGGCTGTACAGTGCAGGCAACCTGCTGCTTGACGCTCAGGAC 16540

Qy     150 TyrArgGlySerValProIleProLeuThr 159
           :::   |||         :::      |||
Db   16541 TTCAGTGGTCAGGGGCAGGTGGTGGCCACC 16570

RESULT 44
US-08-712-072C-1
; Sequence 1, Application US/08712072C
; Patent No. 5925541
;  GENERAL INFORMATION:
;    APPLICANT:  Jack Goldstein, Alex Zhu and Lin Leng
;    TITLE OF INVENTION:  ENDO-BETA-GALACTOSIDASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amster, Rothstein & Ebenstein
;      STREET:  90 Park Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  US
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH 1.44 Mb STORAGE DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/712,072C
;      FILING DATE:  11-SEP-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bogosian, Elizabeth A.
;      REGISTRATION NUMBER:  39,911
;      REFERENCE/DOCKET NUMBER:  63475/97
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 697-5995
;      TELEFAX:  (212) 286-0854 or 286-0082
;      TELEX:  TWX 710-581-4766
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1408 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-712-072C-1

Alignment Scores:
Pred. No.:                37.1            Length:          1408
Score:                    69.00           Matches:         48
Percent Similarity:       33.49%          Conservative:    23
Best Local Similarity:    22.64%          Mismatches:      67
Query Match:              8.14%           Indels:          74
DB:                       2               Gaps:            11

US-09-900-575-29_COPY_26_186 (1-161) x US-08-712-072C-1 (1-1408)

Qy      12 AspLeuSerThrGlnIlePheCysHisAsnAsp------------------------- 22
           ::::|||   |||   |||:::      ||||||
Db    711 AACCTATATACGATGATCTGGAGTCCGAACGACATTCGATTCTACGTCAACAATTCATTG 770

Qy      23 -----------------------------------------TyrProGluThrIleThr 28
                                                    :::|||      :::
Db    771 CAGTATACCTACGCAAGAGTTTCCGGTGGGGGGACACAGCAATGGCCATTTGACGTTCCT 830

Qy      29 AspTyrValThrLeuGlnArg------GlySerAlaTyrGlyGlyValLeuSerAsn--- 45
           |||:::   |||   :::      |||:::   :::   |||   ::::::|||
Db    831 TTTTATCTGATTCTAAATCAGGCCGGTGGAGCGGGATGGCCAGGGGCGATCACAAATGCT 890
```

```
Qy      46 ---------PheSerGlyThrValLysTyr----------------------------- 52
                   ||||||      |||   |||
Db     891 GACTTGCCCTTTAGTATGCAGGTGGATTACGTGCGTGTATATAAGCTGCCTTTATTTAGT 950

Qy      53 --------------SerGlySerSerTyrProPheProThrThrSerGluThrProArg 67
                        ||||||      ||||||:::   |||
Db     951 AATGGCGATTTCGAAAGCGGTGTCATCTATCCATGGACAACATGGGGCGGTGGATCATCG 1010

Qy      68 ValValTyrAsnSer-------------------------ArgThrAspLysProTrp 78
           ||||||                                  |||
Db    1011 GTTGTTTCCACCGATGCCCGGACAGGAACCAAATGCATCCGCGAAACAGGCGGAGAGACA 1070

Qy      79 ProValAlaLeuTyrLeuThrProValSer--------------SerAlaGlyGlyLeu 93
             :::      |||||||||   ::::::                ||||||
Db    1071 TCCATTGAACAATACCTGACCGGTTTAACGCCAAATACGACCTATCGGTTCGGTGGCTAC 1130

Qy      94 ------ValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
                ||| |||||| :::::::::  :::                 |||
Db    1131 GCCAAAGTGTCTGCAGCTGGCCAATCAGTCAGTATTGGTGTC-----------AAAAAT 1178

Qy     112 TyrAsnSerAsp------AspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspVal 129
           |||               |||                 |||:::||||||   |||
Db    1179 TATGGGGGAACTGCGGTCGATGCGACTATAGGTACGACCAGCTACTCCAATAATTCGGTA 1238

Qy     130 ValValProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAsp 149
                  ||||||               |||   :::   |||   |||:::
Db    1239 ACTTTTACAACTGGA--------------GCCAATAATACTACTGCTACGGTCTATTTC 1283

Qy     150 TyrArgGlySerValProIleProLeuThrValTyr 161
           |||:::      |||:::      |||||||||
Db    1284 TATAAA---------CCCTTGAGCGGTACAGTGTAT 1310
```

RESULT 45
US-09-620-412C-308
; Sequence 308, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSEQ for Windows Version 3.0/4.0
; SEQ ID NO 308
; LENGTH: 1860
; TYPE: DNA
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-308

Alignment Scores:

Pred. No.:	55.9	Length:	1860
Score:	69.00	Matches:	35
Percent Similarity:	39.87%	Conservative:	26
Best Local Similarity:	22.88%	Mismatches:	70
Query Match:	8.14%	Indels:	23
DB:	4	Gaps:	7

US-09-900-575-29_COPY_26_186 (1-161) x US-09-620-412C-308 (1-1860)

```
Qy       8 AsnLeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGlu------ 25
           |||:::||||||   :::      |||:::       :::      ||||||
Db    1130 AACATAGTGGTC---GTGGAGGAGCAATTAGCGGTGACTTCGCAGGATCCAGAATTCTTT 1186

Qy      26 ---ThrIleThrAspTyrValThrLeuGlnArgGlySerAla---TyrGlyGlyValLeu 43
                ||||||:::   :::|||   :::      ||||||   :::|||||   :::
Db    1187 TTCTTAATAACCAACAA-ATTACTTTCGAAGGCAATAGCGCTGTGCATGGAGGTGCTATC 1245

Qy      44 SerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro--------- 60
             |||   :::|||   |||:::::::   |||::::::   |||
Db    1246 TACAATAAGAATGGCCTTGTCGAGTTCTTAGGAAATGCAGGACCTCTTGCCTTTAAAGAG 1305
```

```
Qy      61 ---ThrThrSerGluThrProArgValValTyrAsnSerArg---------ThrAspLys 76
                ||||||              :::|||   |||          :::
Db    1306 AACACAACAATAGCTAACGGGGGAGCTATATACACAAGTAATTTCAAAGCGAATCAACAA 1365

Qy      77 ProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLys 96
             |||:::                    :::      ||||||
Db    1366 ACATCCCCCATTCTATTCTCTCAAAATCATGCGAATAAGAAAGGCGGA------------ 1413

Qy      97 AlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAsp 116
                  |||:::                                  |||   |||
Db    1414 --------------GCGATTTACGCGCAATATGTGAACTTAGAACAGAATCAAGATACT 1458

Qy     117 PheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCys 136
           :::|||    |||                          :::   :::      |||
Db    1459 ATTCGCTTTGAAAAAAATACCGCTAAAGAAGGCGGTGGAGCCATCACCTCTTCTCAATGC 1518

Qy     137 AspValSerAlaArgAspValThrValThrLeuProAsp 149
           ::::::|||   :::   |||:::|||      |||
Db    1519 TCAATTACTGCTCATAAT---ACCATCACTTTTTCCGAT 1554
```

Search completed: November 28, 2002, 20:44:20
Job time : 778 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 28, 2002, 19:59:16 ; Search time 53 Seconds
(without alignments)
1169.879 Million cell updates/sec

Title: US-09-900-575-29_COPY_26_186
Perfect score: 848
Sequence: 1 PVVNVGQNLVVDLSTQIFCH.....DVTVTLPDYRGSVPIPLTVY 161

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09900575/runat_22112002_130710_4669/app_query.fasta_1.327
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09900575_@CGN_1_1_21_@runat_22112002_130710_4669
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	848	100.0	837	10	US-09-900-575-7	Sequence 7, Appli
2	841	99.2	837	10	US-09-900-575-17	Sequence 17, Appl
3	834	98.3	837	10	US-09-900-575-6	Sequence 6, Appli
4	834	98.3	837	10	US-09-900-575-21	Sequence 21, Appl
5	834	98.3	840	10	US-09-900-575-10	Sequence 10, Appl
6	834	98.3	840	10	US-09-900-575-12	Sequence 12, Appl
7	834	98.3	903	10	US-09-912-020-207	Sequence 207, App
8	834	98.3	903	10	US-09-900-575-48	Sequence 48, Appl
9	834	98.3	7416	10	US-09-900-575-46	Sequence 46, Appl
10	831	98.0	837	10	US-09-900-575-54	Sequence 54, Appl
11	830	97.9	837	10	US-09-900-575-3	Sequence 3, Appli
12	826	97.4	837	10	US-09-900-575-15	Sequence 15, Appl
13	825	97.3	840	10	US-09-900-575-5	Sequence 5, Appli
14	824	97.2	837	10	US-09-900-575-20	Sequence 20, Appl
15	822	96.9	837	10	US-09-900-575-16	Sequence 16, Appl
16	822	96.9	837	10	US-09-900-575-18	Sequence 18, Appl
17	820	96.7	837	10	US-09-900-575-8	Sequence 8, Appli
18	819	96.6	837	10	US-09-900-575-11	Sequence 11, Appl
19	819	96.6	837	10	US-09-900-575-22	Sequence 22, Appl
20	819	96.6	840	10	US-09-900-575-4	Sequence 4, Appli
21	818	96.5	837	10	US-09-900-575-9	Sequence 9, Appli
22	817	96.3	837	10	US-09-900-575-1	Sequence 1, Appli
23	816	96.2	837	10	US-09-900-575-13	Sequence 13, Appl
24	816	96.2	837	10	US-09-900-575-14	Sequence 14, Appl
25	812	95.8	837	10	US-09-900-575-2	Sequence 2, Appli
26	812	95.8	837	10	US-09-900-575-19	Sequence 19, Appl
27	77.5	9.1	1777	10	US-09-918-568-54	Sequence 54, Appl
c 28	76	9.0	1904	10	US-09-745-763-145	Sequence 145, App
c 29	76	9.0	1977	12	US-10-036-342-79	Sequence 79, Appl
c 30	76	9.0	1977	12	US-10-052-586-503	Sequence 503, App
31	75.5	8.9	1077	10	US-09-815-242-8841	Sequence 8841, Ap
32	75.5	8.9	1346	10	US-09-815-242-4751	Sequence 4751, Ap
33	75.5	8.9	1356	10	US-09-815-242-9029	Sequence 9029, Ap
c 34	74	8.7	17056	10	US-09-893-238-3	Sequence 3, Appli
35	73.5	8.7	4682	10	US-09-783-066-8	Sequence 8, Appli
36	73.5	8.7	17397	10	US-09-764-869-1945	Sequence 1945, Ap
37	73.5	8.7	19334	10	US-09-764-869-1943	Sequence 1943, Ap
38	73.5	8.7	19345	10	US-09-764-869-1944	Sequence 1944, Ap
39	73	8.6	1407	9	US-09-938-842A-2387	Sequence 2387, Ap
40	72.5	8.5	6410	10	US-09-922-217-1094	Sequence 1094, Ap
41	72.5	8.5	17335	10	US-09-764-847-1280	Sequence 1280, Ap
42	72.5	8.5	19882	10	US-09-764-847-1281	Sequence 1281, Ap
43	71.5	8.4	882	10	US-09-925-299-149	Sequence 149, App
44	71.5	8.4	8220	10	US-09-797-908-3	Sequence 3, Appli
c 45	71.5	8.4	9100	10	US-09-452-599-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-09-900-575-7
; Sequence 7, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-7

Alignment Scores:

Pred. No.:	1.09e-104	Length:	837
Score:	848.00	Matches:	161
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0

```
Query Match:            100.00%     Indels:        0
DB:                     10          Gaps:          0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-7 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGTTGGTGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCGTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 2
US-09-900-575-17
; Sequence 17, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
;   LENGTH: 837
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-17

Alignment Scores:
Pred. No.:              9.52e-104   Length:        837
Score:                  841.00      Matches:       159
Percent Similarity:     99.38%      Conservative:  1
Best Local Similarity:  98.76%      Mismatches:    1
Query Match:            99.17%      Indels:        0
DB:                     10          Gaps:          0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-17 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCGTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 3
US-09-900-575-6
; Sequence 6, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
;   LENGTH: 837
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-6

Alignment Scores:
Pred. No.:              8.3e-103    Length:        837
Score:                  834.00      Matches:       158
Percent Similarity:     98.76%      Conservative:  1
Best Local Similarity:  98.14%      Mismatches:    2
Query Match:            98.35%      Indels:        0
DB:                     10          Gaps:          0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-6 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135
```

```
Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGATTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGTGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGTGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 4
US-09-900-575-21
; Sequence 21, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
;   LENGTH: 837
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-21

Alignment Scores:
Pred. No.:              8.3e-103      Length:          837
Score:                  834.00        Matches:         158
Percent Similarity:     98.76%        Conservative:    1
Best Local Similarity:  98.14%        Mismatches:      2
Query Match:            98.35%        Indels:          0
DB:                     10            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-21 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 5
US-09-900-575-10
; Sequence 10, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
;   LENGTH: 840
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-10

Alignment Scores:
Pred. No.:              8.35e-103     Length:          840
Score:                  834.00        Matches:         158
Percent Similarity:     98.76%        Conservative:    1
Best Local Similarity:  98.14%        Mismatches:      2
Query Match:            98.35%        Indels:          0
DB:                     10            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-10 (1-840)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGATTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255
```

```
Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGTGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGTGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

RESULT 6
US-09-900-575-12
; Sequence 12, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 840
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-12

Alignment Scores:			
Pred. No.:	8.35e-103	Length:	840
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-12 (1-840)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::  ||||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

RESULT 7
US-09-912-020-207
; Sequence 207, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001DV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 207
; LENGTH: 903
; TYPE: DNA
; ORGANISM: E. Coli
US-09-912-020-207

Alignment Scores:			
Pred. No.:	9.31e-103	Length:	903
Score:	834.00	Matches:	158
Percent Similarity:	98.76%	Conservative:	1
Best Local Similarity:	98.14%	Mismatches:	2
Query Match:	98.35%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-912-020-207 (1-903)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318
```

```
Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621
```

```
RESULT 8
US-09-900-575-48
; Sequence 48, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
;   LENGTH: 903
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Sequence J96 fimH plus native signal sequence
US-09-900-575-48

Alignment Scores:
Pred. No.:              9.31e-103       Length:         903
Score:                  834.00          Matches:        158
Percent Similarity:     98.76%          Conservative:   1
Best Local Similarity:  98.14%          Mismatches:     2
Query Match:            98.35%          Indels:         0
DB:                     10              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-48 (1-903)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     139 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 198

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     199 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 258

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     259 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 318

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     319 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 378

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     379 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 438

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     439 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 498

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     499 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 558

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     559 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 618

Qy     161 Tyr 161
           |||
Db     619 TAT 621
```

```
RESULT 9
US-09-900-575-46
; Sequence 46, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 46
;   LENGTH: 7416
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Sequence of Plasmid pCGA126-1
US-09-900-575-46

Alignment Scores:
Pred. No.:              2.23e-101       Length:         7416
Score:                  834.00          Matches:        158
Percent Similarity:     98.76%          Conservative:   1
Best Local Similarity:  98.14%          Mismatches:     2
Query Match:            98.35%          Indels:         0
DB:                     10              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-46 (1-7416)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1244 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 1303

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1304 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 1363

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 1423

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1424 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 1483
```

```
Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    1484 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 1543

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1544 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 1603

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1604 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 1663

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    1664 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 1723

Qy     161 Tyr 161
           |||
Db    1724 TAT 1726
```

RESULT 10
US-09-900-575-54
; Sequence 54, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 54
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-54

Alignment Scores:

Pred. No.:	2.1e-102	Length:	837
Score:	831.00	Matches:	157
Percent Similarity:	98.76%	Conservative:	2
Best Local Similarity:	97.52%	Mismatches:	2
Query Match:	98.00%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-54 (1-837)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCATTTCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

RESULT 11
US-09-900-575-3
; Sequence 3, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-3

Alignment Scores:

Pred. No.:	2.86e-102	Length:	837
Score:	830.00	Matches:	157
Percent Similarity:	98.14%	Conservative:	1
Best Local Similarity:	97.52%	Mismatches:	3
Query Match:	97.88%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-3 (1-837)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACCACCAGCGAAACGCCGCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435
```

```
Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 12
US-09-900-575-15
; Sequence 15, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-15

Alignment Scores:

Pred. No.:	9.86e-102	Length:	837
Score:	826.00	Matches:	156
Percent Similarity:	98.14%	Conservative:	2
Best Local Similarity:	96.89%	Mismatches:	3
Query Match:	97.41%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-15 (1-837)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAGAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 13
US-09-900-575-5
; Sequence 5, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 840
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-5

Alignment Scores:

Pred. No.:	1.35e-101	Length:	840
Score:	825.00	Matches:	156
Percent Similarity:	97.52%	Conservative:	1
Best Local Similarity:	96.89%	Mismatches:	4
Query Match:	97.29%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-5 (1-840)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CATGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555
```

```
Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 14
US-09-900-575-20
; Sequence 20, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-20

Alignment Scores:			
Pred. No.:	1.83e-101	Length:	837
Score:	824.00	Matches:	156
Percent Similarity:	97.52%	Conservative:	1
Best Local Similarity:	96.89%	Mismatches:	4
Query Match:	97.17%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-20 (1-837)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGCGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAAAAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTAGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 15
US-09-900-575-16
; Sequence 16, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-16

| Alignment Scores: | | | |
|---|---|---|---|
| Pred. No.: | 3.4e-101 | Length: | 837 |
| Score: | 822.00 | Matches: | 155 |
| Percent Similarity: | 98.14% | Conservative: | 3 |
| Best Local Similarity: | 96.27% | Mismatches: | 3 |
| Query Match: | 96.93% | Indels: | 0 |
| DB: | 10 | Gaps: | 0 |

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-16 (1-837)

```
Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTCATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAGGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATGGCTGTGCTAATTTTGCGACAGACCAATAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCAGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558
```

RESULT 16
US-09-900-575-18
; Sequence 18, Application US/09900575
; Patent No. US20020150587A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
;   LENGTH: 837
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-18

Alignment Scores:
Pred. No.:              3.4e-101      Length:          837
Score:                  822.00        Matches:         155
Percent Similarity:     98.14%        Conservative:    3
Best Local Similarity:  96.27%        Mismatches:      3
Query Match:            96.93%        Indels:          0
DB:                     10            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-18 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||:::|||||||||||||||||||:::|||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 17
US-09-900-575-8
; Sequence 8, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
;   LENGTH: 837
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-8

Alignment Scores:
Pred. No.:              6.31e-101     Length:          837
Score:                  820.00        Matches:         155
Percent Similarity:     96.89%        Conservative:    1
Best Local Similarity:  96.27%        Mismatches:      5
Query Match:            96.70%        Indels:          0
DB:                     10            Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-8 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCCGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAACCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAATTTTTCCGGGACCGTAAAATATAGTGGCAGTAGCTATCCATTTCCG 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGTGGGGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCTACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CATGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCAATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 18
US-09-900-575-11
; Sequence 11, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
```

```
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
;   LENGTH: 837
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-11

Alignment Scores:
Pred. No.:                8.6e-101        Length:         837
Score:                    819.00          Matches:        154
Percent Similarity:       98.14%          Conservative:   4
Best Local Similarity:    95.65%          Mismatches:     3
Query Match:              96.58%          Indels:         0
DB:                       10              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-11 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           |||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||||:::|||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTCGTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 19
US-09-900-575-22
; Sequence 22, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;   LENGTH: 837
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-22

Alignment Scores:
Pred. No.:                8.6e-101        Length:         837
Score:                    819.00          Matches:        154
Percent Similarity:       98.14%          Conservative:   4
Best Local Similarity:    95.65%          Mismatches:     3
Query Match:              96.58%          Indels:         0
DB:                       10              Gaps:           0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-22 (1-837)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           |||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||||:::|||||||||||||||||||:::|||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTCGTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 20
US-09-900-575-4
; Sequence 4, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 840
;   TYPE: DNA
```

; ORGANISM: E. coli
US-09-900-575-4

Alignment Scores:

Pred. No.:	8.64e-101	Length:	840
Score:	819.00	Matches:	154
Percent Similarity:	98.14%	Conservative:	4
Best Local Similarity:	95.65%	Mismatches:	3
Query Match:	96.58%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-4 (1-840)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTGCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||:::|||||||||||||||||||:::||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

RESULT 21
US-09-900-575-9
; Sequence 9, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-9

Alignment Scores:

Pred. No.:	1.17e-100	Length:	837
Score:	818.00	Matches:	154
Percent Similarity:	97.52%	Conservative:	3
Best Local Similarity:	95.65%	Mismatches:	4
Query Match:	96.46%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-9 (1-837)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           ||||||||||:::|||||||||||||||||||:::||||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGCTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTTTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

RESULT 22
US-09-900-575-1
; Sequence 1, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-1

Alignment Scores:

Pred. No.:	1.6e-100	Length:	837
Score:	817.00	Matches:	155
Percent Similarity:	96.89%	Conservative:	1
Best Local Similarity:	96.27%	Mismatches:	5

```
Query Match:          96.34%      Indels:      0
DB:                   10          Gaps:        0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-1 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCCGTCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTATCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGCTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db    196 GGCGTGTTATCTAATTTTTCCGGGATCGTAAAATATAGTGGCAGTAGCTATCCTTTCCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACCACCAGCGAAACGCCGCGCGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCAGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGGTTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGCTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 23
US-09-900-575-13
; Sequence 13, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
;   LENGTH: 837
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-13

Alignment Scores:
Pred. No.:              2.18e-100      Length:          837
Score:                  816.00         Matches:         154
Percent Similarity:     97.52%         Conservative:    3
Best Local Similarity:  95.65%         Mismatches:      4
Query Match:            96.23%         Indels:          0
DB:                     10             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-13 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy     21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy     41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
          |||||||||:::|||||||||||||||||||||:::|||||||||||||||||||||||
Db    196 AGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy     81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
          ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db    316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy    101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy    121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy    141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
          ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db    496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy    161 Tyr 161
          |||
Db    556 TAT 558

RESULT 24
US-09-900-575-14
; Sequence 14, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
;  APPLICANT: Langermann, Solomon
;  APPLICANT:  Revel, Andrew
;  APPLICANT:  Auguste, Christine
;  APPLICANT:  Burlein, Jeanne
;  TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
;  FILE REFERENCE: 469201-549
;  CURRENT APPLICATION NUMBER: US/09/900,575
;  CURRENT FILING DATE:  2001-07-06
;  PRIOR APPLICATION NUMBER: US/60/216,750
;  PRIOR FILING DATE: 2000-07-07
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
;   LENGTH: 837
;   TYPE: DNA
;   ORGANISM: E. coli
US-09-900-575-14

Alignment Scores:
Pred. No.:              2.18e-100      Length:          837
Score:                  816.00         Matches:         154
Percent Similarity:     97.52%         Conservative:    3
Best Local Similarity:  95.65%         Mismatches:      4
Query Match:            96.23%         Indels:          0
DB:                     10             Gaps:            0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-14 (1-837)

Qy      1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
          |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135
```

```
Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           |||||||||:::|||||||||||||||||||||:::|||||||||||||||||||||||
Db     196 AGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

RESULT 25
US-09-900-575-2
; Sequence 2, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-2

Alignment Scores:

Pred. No.:	7.5e-100	Length:	837
Score:	812.00	Matches:	153
Percent Similarity:	96.89%	Conservative:	3
Best Local Similarity:	95.03%	Mismatches:	5
Query Match:	95.75%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-2 (1-837)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTGGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCGGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           |||||||||:::|||||||||||||||||||||:::|||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255

Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||||||||||||||||||||||:::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCCTGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCAGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGCGATGCTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558
```

RESULT 26
US-09-900-575-19
; Sequence 19, Application US/09900575
; Patent No. US20020150587A1
; GENERAL INFORMATION:
; APPLICANT: Langermann, Solomon
; APPLICANT: Revel, Andrew
; APPLICANT: Auguste, Christine
; APPLICANT: Burlein, Jeanne
; TITLE OF INVENTION: FimH Adhesin Proteins and Methods of Use
; FILE REFERENCE: 469201-549
; CURRENT APPLICATION NUMBER: US/09/900,575
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US/60/216,750
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 837
; TYPE: DNA
; ORGANISM: E. coli
US-09-900-575-19

Alignment Scores:

Pred. No.:	7.5e-100	Length:	837
Score:	812.00	Matches:	154
Percent Similarity:	97.52%	Conservative:	3
Best Local Similarity:	95.65%	Mismatches:	4
Query Match:	95.75%	Indels:	0
DB:	10	Gaps:	0

US-09-900-575-29_COPY_26_186 (1-161) x US-09-900-575-19 (1-837)

```
Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHis 20
           |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76 CCTGCCGTGAATGTGGGGCAAAACCTGGTCGTAGATCTTTCGACGCAAATCTTTTGCCAT 135

Qy      21 AsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySerAlaTyrGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     136 AACGATTACCCAGAAACCATTACAGACTATGTCACACTGCAACGAGGTTCGGCTTATGGC 195

Qy      41 GlyValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPhePro 60
           |||||||||:::|||||||||||||||||||||:::|||||||||||||||||||||||
Db     196 GGCGTGTTATCTAGTTTTTCCGGGACCGTAAAATATAATGGCAGTAGCTATCCTTTCCCT 255
```

```
Qy      61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     256 ACTACCAGCGAAACGCCGCGGGTTGTTTATAATTCGAGAACGGATAAGCCGTGGCCGGTG 315

Qy      81 AlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeu 100
           ||||||||||||||   ||||||||||||||||::::   ||||||||||||||||||
Db     316 GCGCTTTATTTGACGCTGGTGAGCAGTGCGGGGGGAGTGGCGATTAAAGCTGGCTCATTA 375

Qy     101 IleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAspPheGlnPheVal 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     376 ATTGCCGTGCTTATTTTGCGACAGACCAACAACTATAACAGCGATGATTTCCAGTTTGTG 435

Qy     121 TrpAsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAla 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     436 TGGAATATTTACGCCAATAATGATGTGGTGGTGCCCACTGGCGGCTGTGATGTTTCTGCT 495

Qy     141 ArgAspValThrValThrLeuProAspTyrArgGlySerValProIleProLeuThrVal 160
           ||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db     496 CGTGATGTCACCGTTACTCTGCCGGACTACCCTGGTTCAGTGCCGATTCCTCTTACCGTT 555

Qy     161 Tyr 161
           |||
Db     556 TAT 558

RESULT 27
US-09-918-568-54
; Sequence 54, Application US/09918568
; Patent No. US20020054882A1
;   GENERAL INFORMATION:
;        APPLICANT: Yoshinobu OKUNO et al.
;        TITLE OF INVENTION: POLYPEPTIDES FOR USE IN GENERATING
;                            ANTI-HUMAN INFLUENZA VIRUS ANTIBODIES (AS AMENDED)
;        NUMBER OF SEQUENCES: 58
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
;             STREET: 2033 K Street, N.W., #800
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20006
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: MS-DOS
;             SOFTWARE: Wordperfect 5.1
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/918,568
;             FILING DATE: 02-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 09/004,422
;             FILING DATE: January 8, 1998
;             APPLICATION NUMBER: 08/443,862
;             FILING DATE: May 22, 1995
;             APPLICATION NUMBER: 08/229,781
;             FILING DATE: April 19, 1994
;             APPLICATION NUMBER: 08/054,016
;             FILING DATE: April 29, 1993
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Warren M. Cheek, Jr.
;             REGISTRATION NUMBER: 33,367
;             REFERENCE/DOCKET NUMBER: <Unknown>
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 202-721-8200
;             TELEFAX: 202-721-8250
;             TELEX: <Unknown>
;    INFORMATION FOR SEQ ID NO: 54:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1777 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;             MOLECULE TYPE: cDNA to genomic RNA
;             HYPOTHETICAL: <Unknown>
;             ANTI-SENSE: <Unknown>
;             FRAGMENT TYPE: <Unknown>
;             ORIGINAL SOURCE:
;                  ORGANISM: A2/Aichi/2/68
;                  STRAIN: <Unknown>
;                  INDIVIDUAL ISOLATE: <Unknown>
;                  DEVELOPMENTAL STAGE: <Unknown>
;                  HAPLOTYPE: <Unknown>
;                  TISSUE TYPE: <Unknown>
;                  CELL TYPE: <Unknown>
;                  CELL LINE: <Unknown>
;                  ORGANELLE: <Unknown>
;             IMMEDIATE SOURCE:
;                  LIBRARY: <Unknown>
;                  CLONE: <Unknown>
;             POSITION IN GENOME:
;                  CHROMOSOME/SEGMENT: <Unknown>
;                  MAP POSITION: <Unknown>
;                  UNITS: <Unknown>
;             FEATURE:
;                  NAME/KEY:
;                  LOCATION:
;                  IDENTIFICATION METHOD:
;                  OTHER INFORMATION:
;             PUBLICATION INFORMATION:
;                  AUTHORS:
;                  TITLE:
;                  JOURNAL:
;                  VOLUME:
;                  ISSUE:
;                  PAGES:
;                  DATE:
;                  DOCUMENT NUMBER:
;                  FILING DATE:
;                  PUBLICATION DATE:
;                  RELEVANT RESIDUES IN SEQ ID NO:
;              SEQUENCE DESCRIPTION: SEQ ID NO: 54:
US-09-918-568-54

Alignment Scores:
Pred. No.:                1.12          Length:          1777
Score:                    77.50         Matches:         44
Percent Similarity:       38.01%        Conservative:    21
Best Local Similarity:    25.73%        Mismatches:      67
Query Match:              9.14%         Indels:          39
DB:                       10            Gaps:            9

US-09-900-575-29_COPY_26_186 (1-161) x US-09-918-568-54 (1-1777)

Qy       1 ProValValAsnValGlyGlnAsnLeuValValAspLeu--------SerThrGlnIle 17
           |||      :::|||   ||||||         ||||||           ::::::
Db     304 CCTCATTGTGATGTTTTTCAAAATGAGACATGGGACCTTTTCGTTGAACGCAGCAAAGCT 363

Qy      18 PheCysHisAsnAspTyrProGluThrIleThrAspTyrValThrLeuGlnArgGlySer 37
           |||      |||   ||||||    :::   ||||||  :::|||:::
Db     364 TTC---AGCAACTGTTACCCTTATGATGTGCCAGATTATGCCTCCCTTAGGTCACTAGTT 420

Qy      38 AlaTyrGlyGlyValLeuSer----------------AsnPheSerGlyThrValLys 51
           |||      |||   |||                   :::::::|||     :::
Db     421 GCCTCGTCAGGCACTCTGGAGTTTATCACTGAGGGTTTCACTTGGACTGGGGTCACTCAG 480

Qy      52 TyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsn 71
           ||||||:::                                |||       :::
Db     481 AATGGGGGAAGCAAT-----------GCTTGCAAAAGGGGACCTGGTAGCGGTTTTTTC 528

Qy      72 SerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGly 91
           ||||||   :::      |||
Db     529 AGTAGACTGAAC------TGG------------------------------------ 543

Qy      92 GlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
             :::   |||:::||||||         ||||||   :::         ||||||
```

```
Db    544 ------TTGACCAAATCAGGAAGCACATATCCAGTGCTGAACGTGACTATGCCAAACAAT 597

Qy    112 TyrAsnSerAspAspPheGlnPheValTrpAsnIleTyr---------AlaAsnAsnAsp 128
Db    598 GACAATTTTGACAAACTA---TACATTTGGGGGATTCACCACCCGAGCACGAACCAAGAA 654

Qy    129 ------------ValValValProThrGlyGlyCysAspValSerAlaArgAspValThr 144
Db    655 CAAACCAGCCTGTATGTTCAAGCATCAGGGAGAGTCACAGTCTCTACCAGGAGAAGCCAG 714

Qy    145 ValThrLeuProAspTyrArgGlySerValPro 155
Db    715 CAAACTATAATCCCGAATATCGGGTCCAGACCC 747

RESULT 28
US-09-745-763-145/c
; Sequence 145, Application US/09745763
; Patent No. US20020065394A1
;  GENERAL INFORMATION:
;       APPLICANT: Jacobs, Kenneth
;                  McCoy, John M.
;                  LaVallie, Edward R.
;                  Collins-Racie, Lisa A.
;                  Evans, Cheryl
;                  Merberg, David
;                  Treacy, Maurice
;                  Spaulding, Vikki
;       TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
;                           ENCODING THEM
;       NUMBER OF SEQUENCES: 219
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genetics Institute, Inc.
;            STREET: 87 CambridgePark Drive
;            CITY: Cambridge
;            STATE: MA
;            COUNTRY: U.S.A.
;            ZIP: 02140
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/745,763
;            FILING DATE: 18-Jun-2000
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Sprunger, Suzanne A.
;            REGISTRATION NUMBER: 41,323
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617) 498-8284
;            TELEFAX: (617) 876-5851
;   INFORMATION FOR SEQ ID NO: 145:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1904 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: double
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 145:
US-09-745-763-145

Alignment Scores:
Pred. No.:                1.98          Length:          1904
Score:                    76.00         Matches:         40
Percent Similarity:       39.10%        Conservative:    21
Best Local Similarity:    25.64%        Mismatches:      44
Query Match:              8.96%         Indels:          51
DB:                       10            Gaps:            7

US-09-900-575-29_COPY_26_186 (1-161) x US-09-745-763-145 (1-1904)

Qy     37 SerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrVal--------------Lys 51
Db    722 TCTGGTAACGGTGGGAAAATGGAGGATACTCAGAGTACGGTGGAGGAGAATACTGCCCGT 663

Qy     52 TyrSerGlySerSerTyrProPheProThrThrSerGlu---------------------- 64
Db    662 CACTCAGGAACAGCTTATAGACTACAAAGGCGATCCCAAGGAGTACCACGATGGTAATCA 603

Qy     65 ---------------------ThrProArgValValTyrAsnSerArgThrAspLysPro 77
Db    602 ATCCACTCATGTTACAGGAATCCGCCGAGGACCACTTATAATAATAATCAGAGAAAGAGG 543

Qy     78 TrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAla 97
Db    542 CAAAGCCGTGCTGCTTTCCAGACTCCTTCAGTTTCTGCAGGCCAAGTTCTGTATAATCTA 483

Qy     98 Gly-----------------SerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
Db    482 AATTATACTCCAAGCCACAAGAACCTCTTAGTACATACTGGTCTTCAGAGGACTCATAGC 423

Qy    112 TyrAsnSerAspAsp------PheGln-PheValTrpAsnIleTyrAlaAsnAsnAspVa 129
Db    422 CTTCACAGCTCACCACAGTTTTTCCAAATTTGTATGCAATATC---------------- 380

Qy    129 lValValProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAs 149
Db    379 ------------------------------------TAAGTCCGTCTTACATTCCCACT 357

Qy    149 pTyrArgGlySerValProIlePro---------LeuThrValTyr 161
Db    356 GTAC------ATCATACCCATCCCAGCCTTTGTTCTGACACTGTAT 317

RESULT 29
US-10-036-342-79/c
; Sequence 79, Application US/10036342
; Patent No. US20020090681A1
; GENERAL INFORMATION:
;  APPLICANT: Desnoyers,Luc
;  APPLICANT:  Eaton,Dan L.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Pan,James
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3030R1C5
;  CURRENT APPLICATION NUMBER: US/10/036,342
;  CURRENT FILING DATE:  2001-12-26
;  PRIOR APPLICATION NUMBER: 60/085579
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/112514
;  PRIOR FILING DATE: 1998-12-15
;  PRIOR APPLICATION NUMBER: 60/113300
;  PRIOR FILING DATE: 1998-12-22
;  PRIOR APPLICATION NUMBER: 60/113430
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/113605
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/113621
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/114140
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/115552
;  PRIOR FILING DATE: 1999-01-12
;  PRIOR APPLICATION NUMBER: 60/116843
;  PRIOR FILING DATE: 1999-01-22
;  PRIOR APPLICATION NUMBER: 60/125774
;  PRIOR FILING DATE: 1999-03-23
;  PRIOR APPLICATION NUMBER: 60/125778
;  PRIOR FILING DATE: 1999-03-23
```

; PRIOR APPLICATION NUMBER: 60/125826
; PRIOR FILING DATE: 1999-03-24
; PRIOR APPLICATION NUMBER: 60/127035
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/127706
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/129122
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/130359
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131272
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/132371
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132379
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132383
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/135750
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/138166
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/144791
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/146970
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/162506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 09/311832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/380142
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/644848
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 09/747259
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/816744
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/854208
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/854280
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/874503
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: 09/869599
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 09/908,827
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US99/10733
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30720
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/05601
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/34956
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 80
; SEQ ID NO 79
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-036-342-79

Alignment Scores:

Pred. No.:	2.1	Length:	1977
Score:	76.00	Matches:	40
Percent Similarity:	39.10%	Conservative:	21
Best Local Similarity:	25.64%	Mismatches:	44
Query Match:	8.96%	Indels:	51
DB:	12	Gaps:	7

US-09-900-575-29_COPY_26_186 (1-161) x US-10-036-342-79 (1-1977)

```
Qy      37 SerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrVal--------------Lys 51
           |||       ||||||   :::   :::          ||||||              :::
Db     830 TCTGGTAACGGTGGGAAAATGGAGGATACTCAGAGTACGGTGGAGGAGAATACTGCCCGT 771

Qy      52 TyrSerGlySerSerTyrProPheProThrThrSerGlu---------------------- 64
           :::|||||:::::|||               |||:::
Db     770 CACTCAGGAACAGCTTATAGACTACAAAGGCGATCCCAAGGAGTACCACGATGGTAATCA 711

Qy      65 ----------------------ThrProArgValValTyrAsnSerArgThrAspLysPro 77
                                   ||||||      |||||||::          |||
Db     710 ATCCACTCATGTTACAGGAATCCGCCGAGGACCACTTATAATAATAATCAGAGAAAGAGG 651

Qy      78 TrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAla 97
                  |||   :::   ||||||      ||||||||   ::::::
Db     650 CAAAGCCGTGCTGCTTTCCAGACTCCTTCAGTTTCTGCAGGCCAAGTTCTGTATAATCTA 591

Qy      98 Gly-----------------SerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
                               :::|||:::           |||:::::::|||::::::
Db     590 AATTATACTCCAAGCCACAAGAACCTCTTAGTACATACTGGTCTTCAGAGGACTCATAGC 531

Qy     112 TyrAsnSerAspAsp------PheGln-PheValTrpAsnIleTyrAlaAsnAsnAspVa 129
             :::|||            |||||| ||||||   ||||||
Db     530 CTTCACAGCTCACCACAGTTTTTCCAAATTTGTATGCAATATC----------------- 488

Qy     129 lValValProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAs 149
                                                  |||   :::|||   |||
Db     487 ----------------------------------TAAGTCCGTCTTACATTCCCACT 465

Qy     149 pTyrArgGlySerValProIlePro---------LeuThrValTyr 161
             |||         :::|||||||||         ||||||:::|||
Db     464 GTAC------ATCATACCCATCCCAGCCTTTGTTCTGACACTGTAT 425
```

RESULT 30
US-10-052-586-503/c
; Sequence 503, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria

; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081838
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082568
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082569
; PRIOR FILING DATE: 1998-04-21
; PRIOR APPLICATION NUMBER: 60/082704
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/082797
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083495
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083496
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083499
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/083559
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084366
; PRIOR FILING DATE: 1998-05-05
; PRIOR APPLICATION NUMBER: 60/084414
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/084639
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084640
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084643
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085580
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085582
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085700
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086023
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086486
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087098
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087208
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05

```
;  PRIOR APPLICATION NUMBER: 60/088212
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088217
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/088326
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088655
;  PRIOR FILING DATE: 1998-06-09
;  PRIOR APPLICATION NUMBER: 60/088722
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088738
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088740
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088811
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088824
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088825
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088826
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088861
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088863
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/088876
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/089090
;  PRIOR FILING DATE: 1998-06-12
;  PRIOR APPLICATION NUMBER: 60/089105
;  PRIOR FILING DATE: 1998-06-12
;  PRIOR APPLICATION NUMBER: 60/089512
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089514
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089538
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089598
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089653
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089908

Alignment Scores:
Pred. No.:                2.1          Length:          1977
Score:                    76.00        Matches:         40
Percent Similarity:       39.10%       Conservative:    21
Best Local Similarity:    25.64%       Mismatches:      44
Query Match:              8.96%        Indels:          51
DB:                       12           Gaps:            7

US-09-900-575-29_COPY_26_186 (1-161) x US-10-052-586-503 (1-1977)

Qy      37 SerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrVal----------Lys 51
           |||     ||||||  :::  :::       ||||||            :::
Db     830 TCTGGTAACGGTGGGAAAATGGAGGATACTCAGAGTACGGTGGAGGAGAATACTGCCCGT 771

Qy      52 TyrSerGlySerSerTyrProPheProThrThrSerGlu-------------------- 64
           :::|||||:::::||          |||:::
Db     770 CACTCAGGAACAGCTTATAGACTACAAAGGCGATCCCAAGGAGTACCACGATGGTAATCA 711

Qy      65 --------------------ThrProArgValValTyrAsnSerArgThrAspLysPro 77
                               ||||||     ||||||:::       |||
Db     710 ATCCACTCATGTTACAGGAATCCGCCGAGGACCACTTATAATAATAATCAGAGAAAGAGG 651

Qy      78 TrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAla 97
               |||  :::  ||||||     |||||||||  ::::::
Db     650 CAAAGCCGTGCTGCTTTCCAGACTCCTTCAGTTTCTGCAGGCCAAGTTCTGTATAATCTA 591

Qy      98 Gly-----------------SerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsn 111
                              :::|||:::           |||::::::|||::::::
Db     590 AATTATACTCCAAGCCACAAGAACCTCTTAGTACATACTGGTCTTCAGAGGACTCATAGC 531

Qy     112 TyrAsnSerAspAsp------PheGln-PheValTrpAsnIleTyrAlaAsnAsnAspVa 129
              :::|||           ||||||  ||||||   ||||||
Db     530 CTTCACAGCTCACCACAGTTTTTCCAAATTTGTATGCAATATC---------------- 488

Qy     129 lValValProThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAs 149
                                                        |||   :::|||   |||
Db     487 ------------------------------------TAAGTCCGTCTTACATTCCCACT 465

Qy     149 pTyrArgGlySerValProIlePro---------LeuThrValTyr 161
             |||         :::|||||||||         ||||||:::|||
Db     464 GTAC------ATCATACCCATCCCAGCCTTTGTTCTGACACTGTAT 425

RESULT 31
US-09-815-242-8841
; Sequence 8841, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8841
;   LENGTH: 1077
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1077)
US-09-815-242-8841

Alignment Scores:
Pred. No.:                0.979        Length:          1077
Score:                    75.50        Matches:         28
Percent Similarity:       41.23%       Conservative:    19
Best Local Similarity:    24.56%       Mismatches:      44
Query Match:              8.90%        Indels:          23
DB:                       10           Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-09-815-242-8841 (1-1077)

Qy      11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr------ 28
           :::::::::::|||:::|||          :::       |||   :::|||
Db     580 ATGGAAATGAATACACGTATTCAAGTAGAACATCCTGTAACTGAAATGGTAACAGGAATT 639

Qy      29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGly 48
           |||   |||   ||||||         |||   |||   ||||||
Db     640 GATTTAGTTAAATTACAATTACAAGTTGCTATGGGTGACGTGTTACCGTATAAACAAGAA 699
```

```
Qy      49 ThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgVal 68
           :::|||   :::|||   :::       |||       :::|||   |||
Db     700 GATATTAAATTAACAGGACACGCAATTGAATTTAGAATTAATGCTGAAAATCCT------ 753

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::       |||       :::       ||||||   |||
Db     754 ---TACAAGAACTTTATGCCATCACCAGGTAAAATTGAGCAATATCTTGCACCA------ 804

Qy      89 SerAlaGlyGlyLeuValIleLysAla------------------------------ 97
           ||||||      ::::::
Db     805 ------GGTGGATATGGTGTTCGAATAGAGTCAGCATGTTATACTAATTATACGATACCG 858

Qy      98 ---------GlySerLeuIleAlaValLeuIleLeuArgGln 108
                   |||::::::|||   ||||||:::   :::
Db     859 CCATATTATGATTCGATGGTAGCGAAATTAATCATACATGAA 900
```

RESULT 32
US-09-815-242-4751
; Sequence 4751, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4751
; LENGTH: 1346
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4751

Alignment Scores:

Pred. No.:	1.37	Length:	1346
Score:	75.50	Matches:	28
Percent Similarity:	41.23%	Conservative:	19
Best Local Similarity:	24.56%	Mismatches:	44
Query Match:	8.90%	Indels:	23
DB:	10	Gaps:	4

US-09-900-575-29_COPY_26_186 (1-161) x US-09-815-242-4751 (1-1346)

```
Qy      11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr------ 28
           :::::::::::|||:::|||       :::       |||   :::|||
Db     861 ATGGAAATGAATACACGTATTCAAGTAGAACATCCTGTAACTGAAATGGTAACAGGAATT 920

Qy      29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGly 48
           |||   |||   ||||||         |||   |||   ||||||
```

```
Db     921 GATTTAGTTAAATTACAATTACAAGTTGCTATGGGTGACGTGTTACCGTATAAACAAGAA 980

Qy      49 ThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgVal 68
           :::|||   :::|||   :::       |||       :::|||   |||
Db     981 GATATTAAATTAACAGGACACGCAATTGAATTTAGAATTAATGCTGAAAATCCT------ 1034

Qy      69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
           |||   :::       |||       :::       ||||||   |||
Db    1035 ---TACAAGAACTTTATGCCATCACCAGGTAAAATTGAGCAATATCTTGCACCA------ 1085

Qy      89 SerAlaGlyGlyLeuValIleLysAla------------------------------ 97
           ||||||      ::::::
Db    1086 ------GGTGGATATGGTGTTCGAATAGAGTCAGCATGTTATACTAATTATACGATACCG 1139

Qy      98 ---------GlySerLeuIleAlaValLeuIleLeuArgGln 108
                   |||::::::|||   ||||||:::   :::
Db    1140 CCATATTATGATTCGATGGTAGCGAAATTAATCATACATGAA 1181
```

RESULT 33
US-09-815-242-9029
; Sequence 9029, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9029
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1356)
US-09-815-242-9029

Alignment Scores:

Pred. No.:	1.39	Length:	1356
Score:	75.50	Matches:	28
Percent Similarity:	41.23%	Conservative:	19
Best Local Similarity:	24.56%	Mismatches:	44
Query Match:	8.90%	Indels:	23
DB:	10	Gaps:	4

US-09-900-575-29_COPY_26_186 (1-161) x US-09-815-242-9029 (1-1356)

```
Qy      11 ValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr------ 28
           :::::::::::|||:::|||       :::       |||   :::|||
```

```
Db    859 ATGGAAATGAATACACGTATTCAAGTAGAACATCCTGTAACTGAAATGGTAACAGGAATT 918

Qy     29 AspTyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGly 48
          |||   |||   ||||||     |||   |||   ||||||
Db    919 GATTTAGTTAAATTACAATTACAAGTTGCTATGGGTGACGTGTTACCGTATAAACAAGAA 978

Qy     49 ThrValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgVal 68
            :::|||   :::|||   :::       |||          :::|||   |||
Db    979 GATATTAAATTAACAGGACACGCAATTGAATTTAGAATTAATGCTGAAAATCCT------ 1032

Qy     69 ValTyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer 88
            |||   :::           |||       :::      ||||||   |||
Db   1033 ---TACAAGAACTTTATGCCATCACCAGGTAAAATTGAGCAATATCTTGCACCA------ 1083

Qy     89 SerAlaGlyGlyLeuValIleLysAla-------------------------------- 97
             ||||||      ::::::
Db   1084 ------GGTGGATATGGTGTTCGAATAGAGTCAGCATGTTATACTAATTATACGATACCG 1137

Qy     98 ---------GlySerLeuIleAlaValLeuIleLeuArgGln 108
                  |||:::::::|||   ||||||:::   :::
Db   1138 CCATATTATGATTCGATGGTAGCGAAATTAATCATACATGAA 1179
```

RESULT 34
US-09-893-238-3/c
; Sequence 3, Application US/09893238
; Patent No. US20020150973A1
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Nagle, D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
; FILE REFERENCE: 7853-237
; CURRENT APPLICATION NUMBER: US/09/893,238
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/245,041
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: 60/093,630
; PRIOR FILING DATE: 1998-07-21
; PRIOR APPLICATION NUMBER: 60/104,978
; PRIOR FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 17056
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-893-238-3

Alignment Scores:

Pred. No.:	101	Length:	17056
Score:	74.00	Matches:	30
Percent Similarity:	43.56%	Conservative:	14
Best Local Similarity:	29.70%	Mismatches:	38
Query Match:	8.73%	Indels:	19
DB:	10	Gaps:	3

US-09-900-575-29_COPY_26_186 (1-161) x US-09-893-238-3 (1-17056)

```
Qy      2 ValValAsnValGlyGlnAsnLeuValValAspLeuSerThrGlnIlePheCysHisAsn 21
          |||:::|||     |||   |||::::::||||||             |||
Db   9975 GTGCTCAATCCTCACCAAGGTCTGATCCTTGACCTTTGAGAATCTTTCTTC--------- 9925

Qy     22 AspTyrProGluThrIleThrAspTyr-ValThrLeuGlnArgGlySerAlaTyrGlyGl 41
                      ::::::   ||| :::||||||:::                 |||||
Db   9924 ---------------CTTTCTACATATTTTGACCCTTGAGGAGCTTATAGTCTTGGGAGG 9880

Qy     41 yValLeuSerAsnPheSerGlyThrValLysTyrSerGlySerSerTyrProPheProTh 61
          |                 |||      :::|||   |||   ||||||
Db   9879 AGAGAGATACACTGATTTGGGGATGACTAGGTATGATGGTTAATCTTATGTAGCTTGATT 9820

Qy     61 rThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProValAl 81
             ||||||||||||||||           :::       ||||||   :::||
Db   9819 GGATTGAGAGACACCTAGAGTT-----------CAAGATGGTATACCTTGGTATATTGC 9772

Qy     81 aLeuTyrLeuThrProValSerSerAlaGlyGlyLeuValIleLysAlaGlySerLeuIl 101
          |                 |||   ||||||||||   :::   :::      :::::
Db   9771 T-----------------AGCTAGGCTGGGGGGCTAGATCTGGCATCAAATCAAGTTGT 9730

Qy    101 e 101
          :
Db   9729 A 9729
```

RESULT 35
US-09-783-066-8
; Sequence 8, Application US/09783066
; Patent No. US20020142302A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Yang, Yonghong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aidong J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20020142302Alel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 791CIP2D
; CURRENT APPLICATION NUMBER: US/09/783,066
; CURRENT FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: No. US20020142302A1 Yet Assigned
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/552,929
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 8
; LENGTH: 4682
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (310)..(4446)
US-09-783-066-8

Alignment Scores:

Pred. No.:	16.7	Length:	4682
Score:	73.50	Matches:	36
Percent Similarity:	40.76%	Conservative:	28
Best Local Similarity:	22.93%	Mismatches:	66
Query Match:	8.67%	Indels:	27
DB:	10	Gaps:	8

US-09-900-575-29_COPY_26_186 (1-161) x US-09-783-066-8 (1-4682)

```
Qy      9 LeuValValAspLeuSerThrGlnIlePheCysHisAsnAspTyrProGluThrIleThr 28
          :::::::::      |||      ::::::                  :::   ::::::
Db   2923 ATTCTGATGGGTTTTAGCATATGGTTGTATTGGCGAAGAAAGAAGAGGAAGGGACTCAGT 2982

Qy     29 AspTyr---ValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSer 47
          ::::|||   ||||||   |||||||||         ||||||::::::||||||   |||
Db   2983 AATTATGCTGTTACGTTTCAAAGAGGAGAT------GGAGGACTAATGAGCAATGGAAGC 3036

Qy     48 ------GlyThrValLysTyrSerGlySerSerTyrPro--------------PhePro 60
                |||   :::          |||||||||                 :::|||
Db   3037 CGTCCAGGTCTTCTCAATGCTGGTGATCCCAGCTATCCATGGCTTGCTGATTCTTGGCCA 3096

Qy     61 ThrThrSerGluThrProArgValValTyrAsnSerArgThrAspLysProTrpProVal 80
             ||||||            :::      |||:::           |||      :::
Db   3097 GCCACGAGC-----------TTGCCAGTAAATAATAGCAACAGTGGCCCAAATGAGATT 3144

Qy     81 AlaLeuTyr--------------LeuThrProValSerSerAlaGly---------Gly 92
                :::               |||   ||||||         |||
```

```
Db   3145 GGAAATTTTGGCCGTGGAGATGTGCTGCCACCAGTTCCAGGCCAAGGGGATAAAACAGCA 3204

Qy     93 LeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyr 112
          ::::::     |||::::::   :::   :::                  :::|||
Db   3205 ACGATGCTCTCAGATGGAGCCATTTATAGTAGCATTGACTTCACTACCAAAACCAGTTAC 3264

Qy    113 AsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValValPro 132
          ||||||                  ||||||         ::::::
Db   3265 AACAGTTCCAGCCAAATAACACAGGCTACCCCATATGCCACGACACAGATCTTGCATTCC 3324

Qy    133 ThrGlyGlyCysAspValSerAlaArgAspValThrValThrLeuProAsp 149
                       |||      ::::::   |||   |||||||||
Db   3325 AAC--------------AGCATACATGAATTGGCTGTCGATCTGCCTGAT 3360

RESULT 36
US-09-764-869-1945
; Sequence 1945, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007
;  CURRENT APPLICATION NUMBER: US/09/764,869
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2442
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1945
;   LENGTH: 17397
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-869-1945

Alignment Scores:
Pred. No.:              121           Length:        17397
Score:                  73.50         Matches:       27
Percent Similarity:     45.45%        Conservative:  23
Best Local Similarity:  24.55%        Mismatches:    45
Query Match:            8.67%         Indels:        15
DB:                     10            Gaps:          5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-764-869-1945 (1-17397)

Qy     34 GlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSer 53
          :::::::|||         |||      :::   ||||||||||||:::::::|||
Db   4034 AAGCAGGGCCCAGAAACTGGGAGCTCAATACAGAATTTTTCGGGCTCCATTAAGGAAGAA 4093

Qy     54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArg 73
          |||::::::::                |||   |||   :::   |||   :::
Db   4094 GGAACAGCATGGTTCTGTGGCCCCAGGAGCGGCACG---CAGCCTGTGCTGGACAAGGGT 4150

Qy     74 ThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeu 93
             :::||||||||||         :::      |||:::::::          |||
Db   4151 TGGTCCCAACCCTGGCCCGAGCAGTCTTTT---CAGCCCATGGCCAAGTGCATGGGCTTT 4207

Qy     94 ValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsn 113
                |||   ||||||   :::   |||:::::::::::::::::|||:::
Db   4208 ---------GCCTGGTCTCTCCAAAGCCCCCTGGTCATTCAGAGGTCCAATCACTGC--- 4255

Qy    114 SerAspAspPheGlnPheValTrpAsnIleTyrAla-----------------AsnAsn 127
                |||         |||:::   :::
Db   4256 ---------TTCCCTGGTTGCTGGAGTCCACACCAGCAGCTCTGGCCTGCTGACACCGAA 4306

Qy    128 AspValValValProThrGlyGlyCysAsp 137
             ||||||            |||  |||
Db   4307 AGTGTGGTGGCTAATGTAGCTGGGACCGAT 4336

RESULT 37
US-09-764-869-1943
; Sequence 1943, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007
;  CURRENT APPLICATION NUMBER: US/09/764,869
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2442
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1943
;   LENGTH: 19334
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-869-1943

Alignment Scores:
Pred. No.:              142           Length:        19334
Score:                  73.50         Matches:       27
Percent Similarity:     45.45%        Conservative:  23
Best Local Similarity:  24.55%        Mismatches:    45
Query Match:            8.67%         Indels:        15
DB:                     10            Gaps:          5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-764-869-1943 (1-19334)

Qy     34 GlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSer 53
          :::::::|||         |||      :::   ||||||||||||:::::::|||
Db   4034 AAGCAGGGCCCAGAAACTGGGAGCTCAATACAGAATTTTTCGGGCTCCATTAAGGAAGAA 4093

Qy     54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArg 73
          |||::::::::                |||   |||   :::   |||   :::
Db   4094 GGAACAGCATGGTTCTGTGGCCCCAGGAGCGGCACG---CAGCCTGTGCTGGACAAGGGT 4150

Qy     74 ThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeu 93
             :::||||||||||         :::      |||:::::::          |||
Db   4151 TGGTCCCAACCCTGGCCCGAGCAGTCTTTT---CAGCCCATGGCCAAGTGCATGGGCTTT 4207

Qy     94 ValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsn 113
                |||   ||||||   :::   |||:::::::::::::::::|||:::
Db   4208 ---------GCCTGGTCTCTCCAAAGCCCCCTGGTCATTCAGAGGTCCAATCACTGC--- 4255

Qy    114 SerAspAspPheGlnPheValTrpAsnIleTyrAla-----------------AsnAsn 127
                |||         |||:::   :::
Db   4256 ---------TTCCCTGGTTGCTGGAGTCCACACCAGCAGCTCTGGCCTGCTGACACCGAA 4306

Qy    128 AspValValValProThrGlyGlyCysAsp 137
             ||||||            |||  |||
Db   4307 AGTGTGGTGGCTAATGTAGCTGGGACCGAT 4336

RESULT 38
US-09-764-869-1944
; Sequence 1944, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007
;  CURRENT APPLICATION NUMBER: US/09/764,869
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2442
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1944
;   LENGTH: 19345
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-869-1944

Alignment Scores:
Pred. No.:              142           Length:        19345
Score:                  73.50         Matches:       27
Percent Similarity:     45.45%        Conservative:  23
Best Local Similarity:  24.55%        Mismatches:    45
Query Match:            8.67%         Indels:        15
DB:                     10            Gaps:          5
```

```
US-09-900-575-29_COPY_26_186 (1-161) x US-09-764-869-1944 (1-19345)

Qy      34 GlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSer 53
           ::::::|||          |||      :::   ||||||||||||:::::|||
Db    4034 AAGCAGGGCCCAGAAACTGGGAGCTCAATACAGAATTTTTCGGGCTCCATTAAGGAAGAA 4093

Qy      54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArg 73
           |||:::::::::          |||   |||   :::   |||   :::
Db    4094 GGAACAGCATGGTTCTGTGGCCCCAGGAGCGGCACG---CAGCCTGTGCTGGACAAGGGT 4150

Qy      74 ThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSerAlaGlyGlyLeu 93
                ::::|||||||||         :::      |||:::::::         |||
Db    4151 TGGTCCCAACCCTGGCCCGAGCAGTCTTTT---CAGCCCATGGCCAAGTGCATGGGCTTT 4207

Qy      94 ValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsnAsnTyrAsn 113
                  |||   ||||||  :::   |||:::::::::::::::|||:::
Db    4208 ---------GCCTGGTCTCTCCAAAGCCCCCTGGTCATTCAGAGGTCCAATCACTGC--- 4255

Qy     114 SerAspAspPheGlnPheValTrpAsnIleTyrAla-----------------AsnAsn 127
                  |||         |||:::   :::
Db    4256 ---------TTCCCTGGTTGCTGGAGTCCACACCAGCAGCTCTGGCCTGCTGACACCGAA 4306

Qy     128 AspValValValProThrGlyGlyCysAsp 137
              ||||||            |||   |||
Db    4307 AGTGTGGTGGCTAATGTAGCTGGGACCGAT 4336

RESULT 39
US-09-938-842A-2387
; Sequence 2387, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2387
;   LENGTH: 1407
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-2387

Alignment Scores:
Pred. No.:              3.18          Length:          1407
Score:                  73.00         Matches:         27
Percent Similarity:     40.18%        Conservative:    18
Best Local Similarity:  24.11%        Mismatches:      27
Query Match:            8.61%         Indels:          40
DB:                     9             Gaps:            5

US-09-900-575-29_COPY_26_186 (1-161) x US-09-938-842A-2387 (1-1407)

Qy      30 TyrValThrLeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThr 49
           :::   :::::::::||||||   ||||||      ||||||:::::::::|||
Db     805 TGGTACTCGATAAAGAGAGGACCGGCTTACATAATCGTGCTAGCTTCATATTCAGCATAT 864

Qy      50 ValLysTyrSerGlySerSerTyrProPheProThrThrSerGluThrProArgValVal 69
             ||||||:::         |||   :::         |||   |||:::|||
Db     865 GGAAAATACACA---CCACAGTACCAATGG---CTCGAAGAGGAGTTCCCAAAGGTT--- 915

Qy      70 TyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSer 89
                ::::||||||:::   ||||||   :::   |||      :::|||
Db     916 ------AACAGAACGGAAACTCCGTGGTTGATTGTTCTGATGCATTCACCATGG------ 963

Qy      90 AlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThr 109
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     963 ---------------------------------------------------------- 963

Qy     110 AsnAsnTyrAsnSerAspAspPheGlnPheVal-------------------------- 120
                ||||||||||   |||:::   ::::::
Db     964 ------TACAACAGCTATGATTACCATTACATGGAAGGTGAAACAATGAGGGTAATGTAT 1017

Qy     121 ------TrpAsnIleTyrAlaAsnAsnAspValVal 130
                   |||   :::            |||||||||
Db    1018 GAGGCTTGGTTTGTCAAGTACAAAGTAGATGTAGTT 1053

RESULT 40
US-09-922-217-1094
; Sequence 1094, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole Lynn
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE:  2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1094
;   LENGTH: 6410
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-217-1094

Alignment Scores:
Pred. No.:              36.5          Length:          6410
Score:                  72.50         Matches:         44
Percent Similarity:     40.62%        Conservative:    21
Best Local Similarity:  27.50%        Mismatches:      59
Query Match:            8.55%         Indels:          37
DB:                     10            Gaps:            9

US-09-900-575-29_COPY_26_186 (1-161) x US-09-922-217-1094 (1-6410)

Qy      24 ProGluThrIleThrAspTyrValThr------LeuGlnArgGlySerAlaTyrGlyGly 41
           |||      :::   :::   ||||||      |||         :::   ||||||
Db     687 CCCCGGGACCTGGCCAACCAGGTGACGGGCTGGCTGGACGGCAGCGCCATCTATGGCTCC 746

Qy      42 -----------------ValLeuSerAsnPheSerGlyThrValLysTyrSerGlySer 55
                                |||   :::|||||||||           ||||||
Db     747 TCGCACTCCTGGAGCGACGCGCTGCGGAGCTTCTCGGGGGGACAGCTGGCGTCGGGGCCC 806

Qy      56 SerTyrProPheProThrThrSerGluThrProArgValValTyr---------AsnSer 72
                ||||||      |||:::   |||   :::::::::         :::
Db     807 GACCCCGCTTTCCCCCGAGACTCGCAGAACCCCCTGCTCATGTGGGCGGCGCCCGACCCC 866

Qy      73 ArgThrAspLysProTrpProValAlaLeuTyrLeu----------------------- 84
              |||   :::      |||      ||||||
Db     867 GCCACCGGGCAGAACGGGCCCCGGGGGCTGTACGCCTTCGGGGCAGAGAGAGGGAACCGG 926

Qy      85 ThrProValSerSerAlaGlyGlyLeuVal---IleLysAlaGlySerLeuIleAlaVal 103
              |||         |||   ||||||:::      :::      :::|||   |||
```

```
Db    927 GAACCCTTCCTGCAGGCGCTGGGCCTGCTCTGGTTCCGCTACCACAACCTGTGGGCGCAG 986

Qy    104 LeuIleLeuArgGlnThrAsnAsnTyrAsnSerAspAsp---PheGlnPheValTrpAsn 122
Db    987 AGGCTGGCCCGCCAGCACCCAGACTGGGAGGACGAGGAGCTGTTCCAG------------ 1034

Qy    123 IleTyrAlaAsnAsnAspValValVal-ProThr-----------------GlyGlyCy 136
Db   1035 ---CACGCACGCAAGAGGGTCATCGCCACCTACCAGAACATCGCTGTGTATGAGTGGCTG 1091

Qy    136 sAspValSerAlaArgAspValThrValThrLeuProAspTyrArgGlySerValPro 155
Db   1092 CCCAGCTTCCTGCAGAAA----------ACACTCCCGGAGTATACAGGATACCGTCCT 1139

RESULT 41
US-09-764-847-1280
; Sequence 1280, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1280
;   LENGTH: 17335
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1280

Alignment Scores:
Pred. No.:                164             Length:         17335
Score:                    72.50           Matches:        34
Percent Similarity:       41.03%          Conservative:   14
Best Local Similarity:    29.06%          Mismatches:     41
Query Match:              8.55%           Indels:         29
DB:                       10              Gaps:           6

US-09-900-575-29_COPY_26_186 (1-161) x US-09-764-847-1280 (1-17335)

Qy     28 ThrAspTyrValThrLeuGln-----------------------------------Arg 35
Db  12243 ACCGATGACGGCACACTACAGTCCCGACTAGGTGAGCTGCCGACGGCGGGGGCTGTGGCC 12302

Qy     36 GlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr--------- 52
Db  12303 GGGTCCCCATGGGGAGGGCTG--------------GGGACTATGCAATTCCAAAGGGAA 12347

Qy     53 ------SerGlySerSerTyrProPheProThrThrSerGlu---ThrProArgValVal 69
Db  12348 TGAAGCGGGGGAGA-TCCTGACCCTACCCTTGTGAGAGTCAAGGCACAGAGTCGGTCTCC 12406

Qy     70 TyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSer 89
Db  12407 CATGAGAGCCGCCCCCCAAACCCCCTCCCTGCACCC---------AGCCCTCAACACCAC 12457

Qy     90 AlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThr 109
Db  12458 CCTCACACACTGGCTCCAGCTGCCAGCTCGTCCGTGACCCTCTTAGTGTCAAAATGGACA 12517

Qy    110 AsnAsnTyrAsnSer------AspAspPheGlnPheValTrpAsnIleTyr 124
Db  12518 AAAGTCTATGGAAGTTTTTAAGCTGATTTTCTGACAAAGTGGCAGGTTTAT 12568

RESULT 42
US-09-764-847-1281
; Sequence 1281, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC009
;  CURRENT APPLICATION NUMBER: US/09/764,847
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2003
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1281
;   LENGTH: 19882
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-847-1281

Alignment Scores:
Pred. No.:                202             Length:         19882
Score:                    72.50           Matches:        34
Percent Similarity:       41.03%          Conservative:   14
Best Local Similarity:    29.06%          Mismatches:     41
Query Match:              8.55%           Indels:         29
DB:                       10              Gaps:           6

US-09-900-575-29_COPY_26_186 (1-161) x US-09-764-847-1281 (1-19882)

Qy     28 ThrAspTyrValThrLeuGln-----------------------------------Arg 35
Db  12251 ACCGATGACGGCACACTACAGTCCCGACTAGGTGAGCTGCCGACGGCGGGGGCTGTGGCC 12310

Qy     36 GlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr--------- 52
Db  12311 GGGTCCCCATGGGGAGGGCTG--------------GGGACTATGCAATTCCAAAGGGAA 12355

Qy     53 ------SerGlySerSerTyrProPheProThrThrSerGlu---ThrProArgValVal 69
Db  12356 TGAAGCGGGGGAGA-TCCTGACCCTACCCTTGTGAGAGTCAAGGCACAGAGTCGGTCTCC 12414

Qy     70 TyrAsnSerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSerSer 89
Db  12415 CATGAGAGCCGCCCCCCAAACCCCCTCCCTGCACCC---------AGCCCTCAACACCAC 12465

Qy     90 AlaGlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThr 109
Db  12466 CCTCACACACTGGCTCCAGCTGCCAGCTCGTCCGTGACCCTCTTAGTGTCAAAATGGACA 12525

Qy    110 AsnAsnTyrAsnSer------AspAspPheGlnPheValTrpAsnIleTyr 124
Db  12526 AAAGTCTATGGAAGTTTTTAAGCTGATTTTCTGACAAAGTGGCAGGTTTAT 12576

RESULT 43
US-09-925-299-149
; Sequence 149, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 149
;   LENGTH: 882
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (14)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (25)
```

```
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (870)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-149

Alignment Scores:
Pred. No.:                2.5          Length:          882
Score:                    71.50        Matches:         31
Percent Similarity:       36.29%       Conservative:    14
Best Local Similarity:    25.00%       Mismatches:      64
Query Match:              8.43%        Indels:          15
DB:                       10           Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-09-925-299-149 (1-882)

Qy     16 GlnIlePheCysHisAsnAspTyrProGluThrIleThrAsp---------TyrValThr 32
          |||   :::|||||||        |||      :::   |||
Db      9 CAAAGNTGGTGTCACCNCGGTGCGCCGTCTAGACTAGTGGATCCCGGCTCAGGATTCGCA 68

Qy     33 LeuGlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyr 52
                 |||   ||||||   ||||||   :::::::|||
Db     69 CGTGCCTCCTGGTCTCAGTATGGCGCTGTCCTGGGTTCTTACAGTCCTGAGCCTCCTACC 128

Qy     53 SerGlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSer 72
          |||      |||   |||   |||        :::              ::::::
Db    129 TCTGCTGGAAGC---CCAGATCCCATTGTGTGCCAACCTAGTACCGGTGCCCATCACCAA 185

Qy     73 ArgThrAspLysPro------TrpProValAlaLeuTyrLeuThrProValSerSerAla 90
          |||         |||      |||   |||   ||||||         :::|||
Db    186 CGCCACCCTGGACCRGATCACTGGCAAGTGGTTTTATATCGCATCGGCCTTTCGAAACGA 245

Qy     91 GlyGlyLeuValIleLysAlaGlySerLeuIleAlaValLeuIleLeuArgGlnThrAsn 110
          |||         :::   :::|||             :::|||:::|||
Db    246 GGAGTACAATAAGTCGGTTCAGGAGATCCAAGCAACCTTCTTTTACTTCACCCCCAACAA 305

Qy    111 AsnTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValVal 130
          :::         |||                                 :::      |||
Db    306 GACAGAGGACACGAT--------------------------CTTTCTCAGAGAGTACCA 338

Qy    131 ValProThrGly 134
             |||||||||
Db    339 GACCCGACAGGA 350

RESULT 44
US-09-797-908-3
; Sequence 3, Application US/09797908
; Patent No. US20020064843A1
; GENERAL INFORMATION:
;  APPLICANT:  WEBSTER, Marion et al.
;  TITLE OF INVENTION:  ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE:  CL00781
;  CURRENT APPLICATION NUMBER: US/09/797,908
;  CURRENT FILING DATE:  2001-03-05
;  NUMBER OF SEQ ID NOS:  9
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 8220
;   TYPE: DNA
;   ORGANISM: Human
US-09-797-908-3

Alignment Scores:
Pred. No.:                72.5         Length:          8220
Score:                    71.50        Matches:         34
Percent Similarity:       42.11%       Conservative:    14
Best Local Similarity:    29.82%       Mismatches:      47
Query Match:              8.43%        Indels:          19
DB:                       10           Gaps:            4

US-09-900-575-29_COPY_26_186 (1-161) x US-09-797-908-3 (1-8220)

Qy     36 GlySerAlaTyrGlyGlyVal-LeuSerAsnPheSerGlyThrValLysTyrSerGlySe 55
          ||||||||||::      ::: ||||||      |||           :::::|||
Db   4047 GGCTCTGCTCACAACTCCATGGCTTTCCTTCCTCTCTACCTTGTGCCCTCATAATGGTTT 4106

Qy     55 rSerTyrProPheProThrThrSerGluThrProArgValValTyrAsnSerArgThrAs 75
                |||   |||   :::|||      |||
Db   4107 CTACCTCCCACTTCCTCTGTCCTCATCTTTACCC----------------------TC 4142

Qy     75 pLysProTrpProVal-Ala------LeuTyrLeuThrProValSerSerAlaGlyGlyL 93
          |||||||||    |||      |||  |||:::|||:::   |||
Db   4143 TGACCCCTGGCCCTTCAGCTCCCAGTCTAAAACTAAGCCCTCTCCCCAGCCTGATGCCCA 4202

Qy     93 euValIleLysAlaGlySerLeuIleAlaVal------LeuIleLeuArgGlnThrAsnA 111
            :::      ||||||      :::   :::      |||:::|||   :::      |
Db   4203 GCCTTTCTGCTGCTGGTAGGGACTTAAAGTTGGAGAACCTGTTGCTGGACAAGTGGGAGA 4262

Qy    111 snTyrAsnSerAspAspPheGlnPheValTrpAsnIleTyrAlaAsnAsnAspValValV 131
          ||                ||||||      |||         :::   :::|||      |||
Db   4263 ATGTGAAGATATCAGACTTTGGCTTTGCCAAGATGGTGCCTTCTAACCAGCCTGTG---- 4318

Qy    131 alProThrGlyGlyCysAspValSerAlaArgAspVal 143
                  ||||||      |||   |||   |||
Db   4319 ----------GGTTGTAGCCCTTCTTACCGCCAAGTG 4345

RESULT 45
US-09-452-599-27/c
; Sequence 27, Application US/09452599
; Patent No. US20020055101A1
; GENERAL INFORMATION:
;  APPLICANT:  Bergeron, Michel G.
;  APPLICANT:  Ouellette, Marc
;  APPLICANT:  Roy, Paul H.
;  TITLE OF INVENTION:  Specific and Universal Probes and Amplification Primers
;  TITLE OF INVENTION:  to Rapidly Detect and Identify Common Bacterial
;  TITLE OF INVENTION:  Pathogens and Antibiotic Resistance Genes from Clinical
;  TITLE OF INVENTION:  Specimens for Routine Diagnosis in Micro
;  FILE REFERENCE:  12287.31
;  CURRENT APPLICATION NUMBER: US/09/452,599
;  CURRENT FILING DATE:  1999-12-01
;  PRIOR APPLICATION NUMBER:  08/526,840
;  PRIOR FILING DATE:  1995-09-11
;  PRIOR APPLICATION NUMBER:  08/304,732
;  PRIOR FILING DATE:  1994-09-12
;  NUMBER OF SEQ ID NOS:  177
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 27
;   LENGTH: 9100
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
US-09-452-599-27

Alignment Scores:
Pred. No.:                84.5         Length:          9100
Score:                    71.50        Matches:         39
Percent Similarity:       44.53%       Conservative:    22
Best Local Similarity:    28.47%       Mismatches:      57
Query Match:              8.43%        Indels:          19
DB:                       10           Gaps:            8

US-09-900-575-29_COPY_26_186 (1-161) x US-09-452-599-27 (1-9100)

Qy     34 GlnArgGlySerAlaTyrGlyGlyValLeuSerAsnPheSerGlyThrValLysTyrSer 53
          :::::|||      |||        |||        ||||||      :::|||
Db    884 GAAAAAGGCTTTTTTTAT---------TTGGATACGCACTCTGGTGTGGGGCGTTATCGT 834

Qy     54 GlySerSerTyrProPheProThrThrSerGluThrProArgValValTyrAsn------ 71
          ||||||               |||   |||           :::
Db    833 TTATCCTCAAATGAATCAGAAAAAACGGGGGAATATAAAGAAGGTATTGGACGCCTGTGG 774

Qy     72 SerArgThrAspLysProTrpProValAlaLeuTyrLeuThrProValSer------Ser 89
          :::|||||||  |||      :::|||   |||:::      ::: .        :::
Db    773 GATCAAACAGATTTACCCGAAGATATTGCTCGTTATGTAAAAATGATCAAAAAACTCAAT 714
```

```
Qy      90 AlaGlyGlyLeuValIleLys------AlaGlySer---LeuIleAlaValLeuIleLeu 106
           ||||||      ::::::      |||||||||   |||||||||      :::|||
Db     713 TATGGTGGCAAAGAACTACGTTATTACGCGGGTTCTCCATTAATTGCCGCGGAATTGTTG 654

Qy     107 ArgGlnThrAsn--------------AsnTyrAsnSerAspAspPheGlnPheValTrp 121
           |||       :::                     :::       |||:::       :::
Db     653 CGCTCACAAGATCGCGCACTATTGACCGAGCTTCATCCTAGCGATTATCCAATTCTTCGC 594

Qy     122 AsnIleTyrAlaAsnAsnAspValValValProThrGlyGlyCysAspValSerAlaArg 141
           |||   ::::::::::::       |||      |||      ||||||          :::
Db     593 AATAATTTTAGCGACGACAAAAATGTC------ACCGTAAAATGTGACAATGGCTTTCAA 540

Qy     142 AspValThrValThrLeuPro------AspTyrArgGlySerValProIle 156
           |||      |||||||||      :::   ||||||   |||   |||
Db     539 CAAGTCAAAGCAACGCTTCCGCCAAAAGAACGCCGAGGCTTAGTACTCATC 489

Search completed: November 28, 2002, 21:29:40
Job time : 80 secs
```